EPAT
Emergency Physicians Association of Turkey

EAJEM

Eurasian Journal of Emergency Medicine

Citation abbreviation: Eurasian J Emerg Med

ISSN 2149-5807 • EISSN 2149-6048

Volume: 25

www.eajem.com

2026

galenos yayınevi

Official Journal of Emergency Physicians Association of Turkey

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Letter to the Editor
DOI: 10.4274/eajem.galenos.2025.88964
Eurasian J Emerg Med. 2026;25: 1

Investigation of the Efficacy of Risk Scoring Systems on Prognosis in Patients with STEMI Presenting to The Emergency Department

Erkan Boğa

Esenyurt Necmi Kadıoğlu State Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye

Keywords: ST-elevation myocardial infarction, risk scoring systems, triage in emergency departments early warning score

Dear Editor,

I read with great interest the article titled "investigation of the efficacy of risk scoring systems on prognosis in patients with ST-elevation myocardial infarction (STEMI) presenting to the emergency department," published in volume 23 of the Eurasian Journal of Emergency Medicine in 2024 (1). This study stands out for its evaluation of the effectiveness of various risk scoring systems used in emergency departments for patients with STEMI.

The findings highlight that the Triage in Emergency Departments Early Warning score (TREWS) performs as effectively as other widely used scoring systems (TIMI, ProACS, and C-ACS) in predicting short-term mortality, even demonstrating superior performance in certain instances. Particularly the high area under the curve values for TREWS (0.847 and 0.823 for 24-hour and 28-day mortality, respectively) emphasize its potential utility in clinical emergency medicine.

However, I would like to draw attention to a few points:

Study design: The single-center design and relatively small sample size may limit the generalizability of the findings. A larger, multicenter study could help validate these results more robustly (2).

Long-term outcomes: While the study focuses on short-term mortality, the evaluation of long-term outcomes (e.g., recurrent cardiac events or quality of life) would provide valuable insights for clinicians making prognostic decisions (3).

Practicality of TREWS: Additional discussion on the practical ease of using TREWS compared to other scoring systems in the time-sensitive setting of emergency departments could further highlight its clinical relevance and applicability (4).

This study provides critical insights and valuable guidance for the management of STEMI patients in emergency settings. The robust statistical analysis and methodology lend credibility to the findings. Addressing the points above in future studies may enhance the understanding and application of these risk scoring systems.

Footnotes

Conflict of Interest: No conflict of interest was declared by the author.

Financial Disclosure: The author declared that this study received no financial support.

References

1. Keski C, Yüksel M, Şensoy B, Ay MO, Avcu N, Kılıç A, et al. Investigation of the efficacy of risk scoring systems on prognosis in patients with STEMI presenting to the emergency department. Eurasian J Emerg Med. 2024;23:242-50.
2. Smith P, Johnson R, Brown T, Davis M, Miller J, Wilson A, et al. Multicenter validation of risk scores in acute coronary syndromes: implications for clinical practice. J Am Coll Cardiol. 2020;75:1456-64.
3. Anderson JL, Adams CD, Antman EM, Holmes Jr DR, Jaffe AS, Jneid H, et al. ACC/AHA guidelines for the management of patients with acute coronary syndromes: a report of the American College of Cardiology/American Heart Association Task Force. J Am Coll Cardiol. 2022;80:e223-e293.
4. Huang J, Wei X, Wang Y, Li Y, Zhang H, Liu Y, Prognostic accuracy of TIMI, GRACE, and C-ACS scores in acute coronary syndromes: a meta-analysis. Eur Heart J Acute Cardiovasc Care. 2021;10:256-64.

Corresponding Author: Erkan Boğa MD, Esenyurt Necmi Kadıoğlu State Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye
E-mail: drerkanboga@gmail.com **ORCID ID:** orcid.org/0000-0001-6802-6301

Cite this article as: Boğa E. Investigation of the efficacy of risk scoring systems on prognosis in patients with STEMI presenting to the emergency department. Eurasian J Emerg Med. 2026;25: 1.

Received: 27.01.2025
Accepted: 24.03.2025
Epub: 21.05.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.93709
Eurasian J Emerg Med. 2026;25: 2-7

Investigation of Associated Diseases in Acute Myocardial Infarction and Heart Failure: Metabolomic Insights

Yasin Ali Çimen[1], Bahadır Taşlıdere[2], Ufuk Sarıkaya[3], Metin Demirel[3], Sahabettin Selek[3]

[1]Bezmialem Vakıf University Faculty of Medicine, Department of Physiology, İstanbul, Türkiye
[2]Bezmialem Vakıf University Faculty of Medicine, Department of Emergency Medicine, İstanbul, Türkiye
[3]Bezmialem Vakıf University Faculty of Medicine, Department of Medical Biochemistry, İstanbul, Türkiye

Abstract

Aim: Cardiovascular diseases (CVDs) are among the most common causes of death world-wide. Acute myocardial infarction (AMI) and heart failure (HF) are the most common CVDs, and the development of HF after AMI is common. The aim of our study was to determine the metabolic profile in the serum of patients with AMI/HF and to correlate metabolites with comorbidities.

Materials and Methods: The study included 30 AMI and 30 HF patients admitted to Emergency Medicine Service who met the inclusion criteria. Serum samples were collected and analysed using liquid chromatography-high resolution mass spectrometry (LC-HRMS). Data were analysed using TidyMass and MetaboAnalyst.

Results: LC-HRMS analysis of AMI and HF revealed metabolites that may be directly associated with pancreatic cancer, colon cancer, renal failure and uremia. These metabolites were involved in energy metabolism, lipid metabolism, and nucleotide metabolism. In addition, the metabolite N,N-dimethyldodecylamine N-oxide was found to be elevated in HF, compared to AMI.

Conclusion: This study sheds light on metabolic changes in AMI/HF patients. It particularly highlights the metabolite relationship between cancer and kidney pathologies and CVDs. It points out that polypharmacy observed in HF may increase the accumulation of possibly hamful chemicals in the body.

Keywords: Acute myocardial infarction, cardiovascular diseases, comorbidities, heart failure, metabolomics, metabolites

Introduction

Cardiovascular diseases (CVD) are one of the leading causes of death world-wide despite advances in diagnosis and treatment (1,2). Acute myocardial infarction (AMI) and heart failure (HF) affect quality of life and impose significant economic burdens (3,4). Although the diagnosis and treatment of AMI patients have improved significantly in the last decade, AMI still remains the most important contributor to HF (5). The incidence of HF among patients hospitalised for AMI ranges from 14% to 36% (6). There are also studies reporting that this rate rises to 37.5% (7). The high mortality rate has forced researchers to look for the best way to diagnose, risk stratify, and manage patients with suspected CVDs. The most frequently studied conditions are AMI and HF (8). Therefore, circulatory biomarkers that provide easy and fast results are utilised in the diagnosis and treatment of these diseases. The most commonly used biomarkers in the diagnosis of AMI and HF are troponin T/I and N-terminal pro-brain natriuretic peptide (9,10). In addition to these classical biomarkers, the identification of new biomarkers with technological advances may facilitate the management of AMI and HF. It may also facilitate the determination of HF development after AMI.

The metabolomics method is a powerful technique that enables the determination of the metabolite profile associated with the disease state (11,12). Metabolite profiling allows a comprehensive assessment of metabolites that may change during disease progression (12,13). Therefore, it can be used to recognize altered metabolic features in diseases and to predict comorbidities. The prediction of the hospitalisation duration and survival probability

Corresponding Author: Yasin Ali Çimen PhD, Bezmialem Vakıf University Faculty of Medicine, Department of Physiology, İstanbul, Türkiye
E-mail: yasin.alicimen@gmail.com **ORCID ID:** orcid.org/0000-0002-8787-3108
Cite this article as: Çimen YA, Taşlıdere B, Sarıkaya U, Demirel M, Selek S. Investigation of associated diseases in acute myocardial infarction and heart failure: metabolomic insights. Eurasian J Emerg Med. 2026;25: 2-7.

Received: 12.03.2025
Accepted: 10.04.2025
Epub: 02.05.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

of patients with CVD by metabolic studies highlights the importance of these studies (14). In addition, serum metabolites have been reported to improve new AMI or HF-specific risk prediction and identification of risk populations (14-16). Despite the abundance of metabolomic studies focusing individually on AMI or HF, comparative analyses between conditions are still scarce. We hypothesized that distinct metabolic profiles may characterize AMI and HF, and that specific metabolites could serve as biomarkers for HF development following AMI. Moreover, we proposed that serum metabolite signatures might aid in identifying comorbidities associated with these conditions. To investigate these hypotheses, we conducted comprehensive non-targeted metabolomic profiling of serum samples from AMI and HF patients.

Materials and Methods

The study included 30 AMI and 30 HF patients admitted to Bezmialem Vakıf University Hospital Emergency Medicine Service. Signatures were obtained from all patients for the informed consent form. Metabolomics was performed on blood samples obtained during routine examination. The AMI patient group consisted of individuals who presented to the emergency department with dyspnoea and chest pain, and were diagnosed with acute coronary artery disease after coronary angiography with more than 50% stenosis. The HF patient group consisted of individuals who presented to the emergency department with shortness of breath and chest pain, and were diagnosed with HF based on physical examination and B-type natriuretic peptide results.

Blood samples in tubes without separating gel were centrifuged at 3500 RPM for 10 minutes. Supernatants were stored at -86 °C until further processing. On the experimental day, 500 µL of thawed serum samples were transferred into a separate Eppendorf tube, followed by the addition of 1 mL of methanol. The mixture underwent centrifugation at 10,000 × g for 1 hour, after which the upper phase was carefully transferred into high-performance liquid chromatography vials. Mass spectrometry analyses were carried out using liquid chromatography, high resolution mass spectrometry on a Thermo Q Exactive instrument. Chromatographic separation was achieved under isocratic conditions with methanol as the mobile phase, employing a Fortis C18 column (3 µm particle size, 150 × 2 mm). Full-scan analysis in both positive and negative electrospray ionization modes was performed. The acquired raw mass spectrometry data were converted into mzXML and mgf formats using the open-source ProteoWizard software (17). The study was conducted in accordance with the principles of the 1964 Declaration of Helsinki. Ethics committee approval was obtained from Bezmialem Vakıf University Non-Interventional Research Ethics Committee (decision number: 2023/312, date: 22.11.2023).

Statistical Analysis

The processed data were analyzed for metabolite profiling using the TidyMass and MetaboAnalyst R packages (18,19). Metabolite annotation was based on publicly available spectral libraries, including the Human Metabolome Database, MassBank, and the North American MassBank. Pathway analysis was conducted using the Kyoto Encyclopedia of Genes and Genomes database. Differential metabolite analysis was conducted using fold change (FC) values, with a significance threshold of FC ≥Q2 or FC ≤0.5. Statistical significance was assessed through adjusted p-values (p-adjust <0.05) using the Benjamini-Hochberg correction to control for false discovery rates. Non-parametric tests, such as the Wilcoxon rank-sum test, were applied to account for non-normally distributed data. Analyses were performed using the open-source coding language R (version 4.4.0) (20).

Results

The mean age and gender of 30 AMI and 30 HF patients included in the study is presented (Table 1). Among the identified metabolite sets, those associated with pancreatic cancer exhibited the most significant enrichment, with a p value of 0.00098. This was followed by kidney disease-associated metabolites, which showed notable enrichment with a p value of 0.00131. Colorectal cancer-associated metabolites also demonstrated significant enrichment, with a p value of 0.00252. The uremia-associated metabolites had the highest p value among the identified sets (p=0.00430), indicating relatively less statistical significance compared to the other enriched pathways (Figure 1).

In this study, a set of circulating metabolites was identified in patients with AMI and HF. The detected metabolites included amino acids and their derivatives such as L-histidine, L-carnitine, DL-phenylalanine, arginine, tryptophan, and norleucine; as well as metabolites related to lipid metabolism, including sphingosine C16, glycerophosphocholine, and lauric acid diethanolamide. Additionally, compounds involved in nucleotide metabolism, such as cytosine, 5-methylcytosine, thymine, and thymidine, were detected. Energy metabolism-related metabolites, including creatinine and creatine, were also identified.

Quantitative comparison of the two patient groups demonstrated statistically significant differences in specific serum metabolites. Among these, N,N-Dimethyldodecylamine N-oxide exhibited the greatest increase in HF patients relative to those with AMI, with a 77.32-FC (p-adjust =0.0018). Conversely, the levels of Ketamine (FC =0.019, p-adjust =0.027) and its metabolite Norketamine (FC =0.035, p-adjust =0.027) were markedly reduced in the HF group

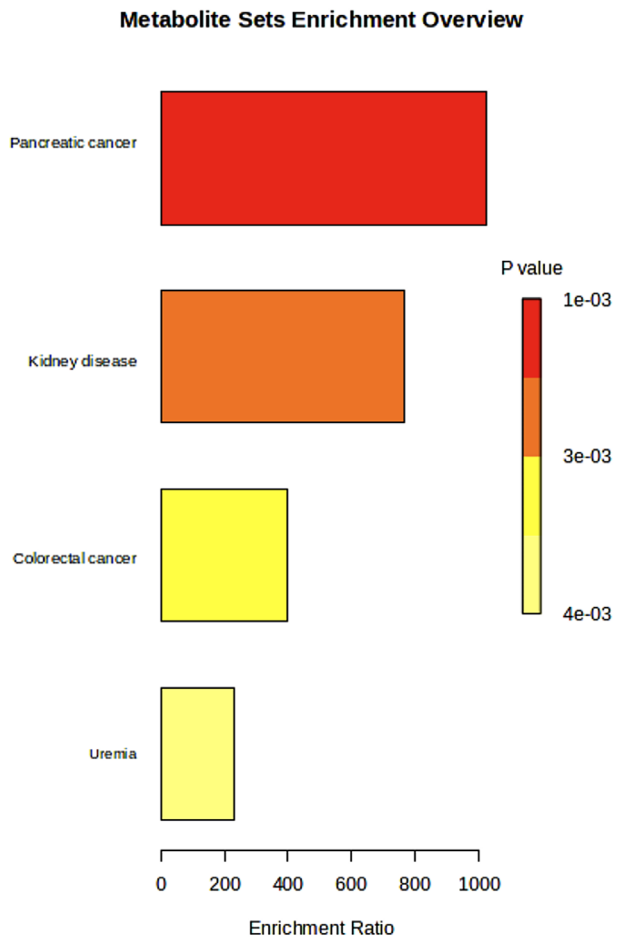


Figure 1. Metabolite set enrichment analysis

This figure presents the metabolite set enrichment analysis conducted to determine the biological relevance of identified metabolites in patients with AMI and HF. The X-axis represents the enrichment ratio, while the color scale indicates the p-value, with lower p-values signifying more significant findings. AMI: Acute myocardial infarction, HF: Heart failure

Table 1. Demographic and clinical characteristics of participants

		AMI	HF
Age	mean ± SD	59.93±14.67	65.87±12.46
	min-max	37-88	34-88
Gender		22M, 8FM	20M, 10FM

AMI: Acute myocardial infarction, HF: Heart failure, FM: Female, M: Male

(Table 2 and Figure 2).

Discussion

The important findings of our study are that common metabolites were detected in pancreatic cancer, kidney disease, colorectal cancer, and uremia, and that these may be associated with AMI-HF. In addition, the N, N-Dimethyldodecylamine N-oxide metabolite was increased in HF compared to AMI, as determined by the examination of AMI and HF metabolic profiles.

AMI and cancer have been reported to involve the same molecular pathways in disease development and progression (5,21). HF is accompanied by a broad spectrum of both cardiovascular and non-cardiovascular comorbidities (22). Recently, the relationship between HF and cancer has been questioned in many studies. In these studies, it has been reported that HF and cancer frequently overlap and evidence of a direct effect between the diseases has emerged (23-25). It has been reported that glucose, glutamine, and fatty acids can provide nutrients to meet the metabolic needs of the tumour (26). In our study of metabolic profiles, we found that metabolites detected in AMI and HF may be associated with pancreatic cancer and colorectal cancer. In particular, common nucleotide and energy metabolites such as cytosine, 5-methylcytosine, thymine, and thymidine strengthened the association between HF and both pancreatic and colorectal cancer. In light of the reported changes in cancer-related energy and nucleotide metabolism (27), our data support these metabolic trends.

The CVD mortality rate increases with decreasing estimated glomerular filtration rate (28). Renal failure is often associated with HF and worsens the patient's prognosis (29). Accurate estimation of renal function may be critical for assessing prognosis in patients with HF (30). In our study, we identified metabolites that are common in nucleotide, energy, and lipid metabolism in AMI-HF and renal disease. A large number of lipid abnormalities have been reported in patients with chronic renal failure, supporting our study results (31). In this respect, the common metabolites detected in our study may be useful in the detection of renal diseases that may accompany CVD.

Previous studies have reported various metabolites such as 9-cis retinoic acid and dehydrophytosphingosine as CVD risk biomarkers (32-34). A recent study found that serum acylcarnitine was associated with clinical symptoms and severity of coronary artery disease (35). These findings underscore the importance of metabolic alterations in the pathophysiology of CVDs. In parallel with technological advancements, the large-scale production and widespread use of organic chemicals in industry and manufacturing have significantly increased (36). In Japan, N,N-dimethyldodecylamine is included in "Class I Designated Chemical Substances", which includes compounds that may be harmful to health and ecosystems (37-39). Alkyldimethylamines, especially N,N-Dimethyldodecylamine, are widely used in many industrial products such as disinfectants, detergents, dye additives, wetting agents, antistatic agents, and textile bleaches (40,41). One study mentioned that dimethyldodecylamine may also be found in medicines (38). Polypharmacy occurs in HF

Table 2. Fold change, adjusted p-value, p-value values for annotated metabolites obtained from analysis results				
Compound name	**Fold change**	**p value**	**p value adjust**	**Adduct**
(+/-)-Norketamine	0.035205	0.00245	0.02704	(M+H)+
N,N-Dimethyldodecylamine N-oxide	77.32334	0.00005	0.00182	(M+H)+
Ketamine	0.019157	0.00188	0.02704	(M+H)+

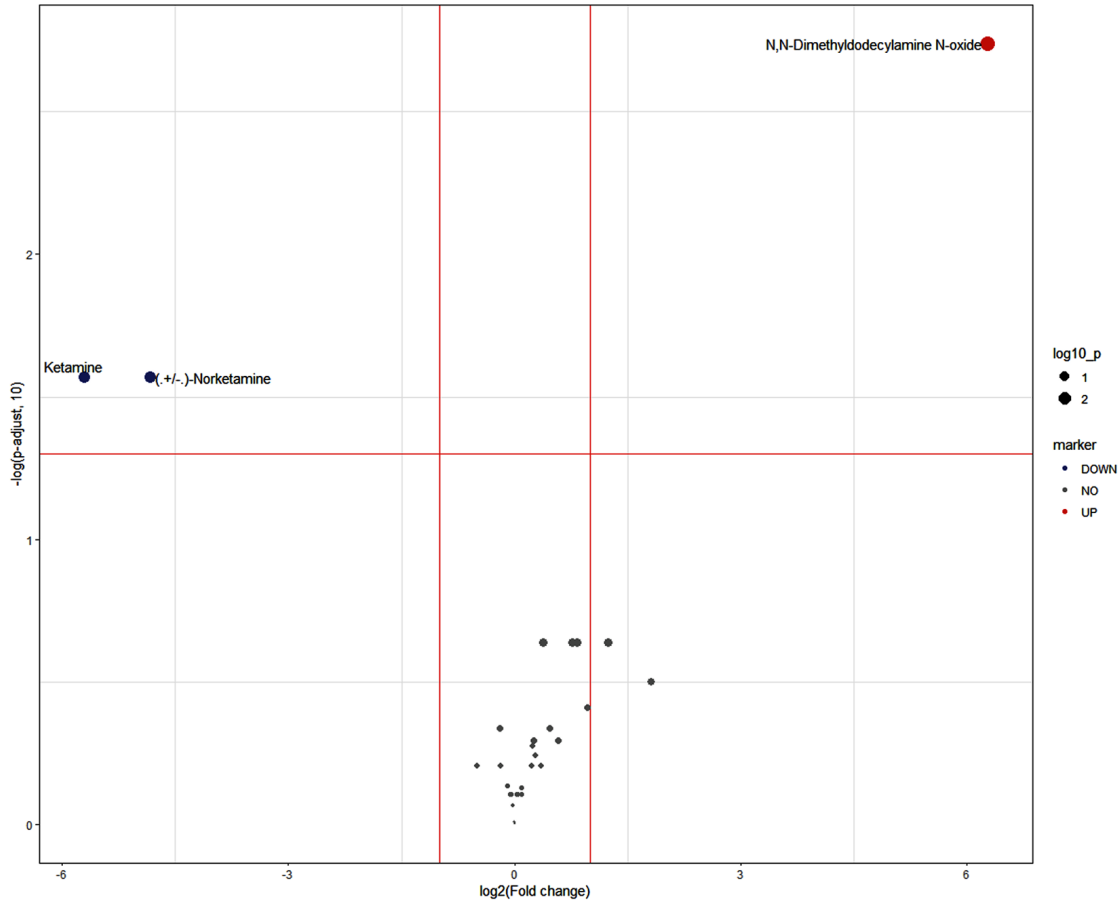


Figure 2. Volcano plot of circulating metabolites in AMI and HF patients

This volcano plot illustrates the comparative analysis of circulating metabolite levels in HF patients relative to those with AMI. The X-axis represents $\log_2$ (fold change) values, while the Y-axis denotes $-\log_{10}$ (p-adj) values. Statistically significant metabolites are color-coded: red indicates upregulated (UP) metabolites, while blue represents downregulated (DOWN) metabolites. AMI: Acute myocardial infarction, HF: Heart failure

because there are many comorbidities (42). Polypharmacy is the long-term use of 5 or more drugs (43). Dimethyldodecylamine N-oxide has been reported to have a median concentration of 29.9 ng/mL in human serum (44). In our study, we found that the concomitant use of various drugs may have caused an increase in N,N-dimethyldodecylamine N-oxide metabolite in HF. In addition, we found that serum ketamine and norcetamine metabolites were reduced in HF, compared to AMI. This result was not very plausible. This unexpected result may be because our study was a non-targeted metabolomics study.

Study Limitations

The small number of patients participating in the study, the fact that the study was conducted in a single hospital, and the absence of a control group limit our study.

Conclusion

Our results showed that nucleotide, energy, and lipid metabolites are common in AMI/HF, cancer, and kidney diseases. Detection of these metabolites may accelerate the detection and treatment of AMI and HF comorbidities. Thus, the management of AMI

and HF may become easier. In addition, excessive drug use in HF may cause harmful chemicals to accumulate in the body and complicate the treatment process.

Ethics

Ethics Committee Approval: The study was conducted in accordance with the principles of the 1964 Declaration of Helsinki. Ethics committee approval was obtained from Bezmialem Vakıf University Non-Interventional Research Ethics Committee (decision number: 2023/312, date: 22.11.2023).

Informed Consent: Signatures were obtained from all patients for the informed consent form.

Footnotes

Author Contributions

Concept: Y.A.Ç., B.T., U.S., M.D., S.S., Design: Y.A.Ç., B.T., U.S., M.D., S.S., Data Collection or Processing: Y.A.Ç., B.T., Analysis or Interpretation: U.S., M.D., S.S., Literature Search: A.Ç., B.T., Writing: Y.A.Ç., M.D., S.S.

Conflict of Interest: The authors declare that they have no conflict of interest.

Financial Disclosure: Financial Disclosure: This study was supported by Bezmialem Vakıf University Scientific Research Projects Coordination Office (project number: 20231201).

References

1. Wilkins JT, Ning H, Berry J, Zhao L, Dyer AR, Lloyd-Jones DM. Lifetime risk and years lived free of total cardiovascular disease. JAMA. 2012;308:1795-1801.
2. Roth GA, Huffman MD, Moran AE, Feigin V, Mensah GA, Naghavi M, et al. Global and regional patterns in cardiovascular mortality from 1990 to 2013. Circulation. 2015;132:1667-78.
3. Asano T, Ono M, Dai Z, Saito A, Kanie T, Takaoka Y, et al. T Temporal trends in clinical outcomes after percutaneous coronary intervention: a systematic review of 66,327 patients from 25 all-comers trials. EuroIntervention. 2022;17:1318-29.
4. Caidahl K, Hartford M, Ravn-Fischer A, Lorentzen E, Yndestad A, Karlsson T, et al. Homeostatic chemokines and prognosis in patients with acute coronary syndromes. J Am Coll Cardiol. 2019;74:774-82.
5. Ibanez B, James S, Agewall S, Antunes MJ, Bucciarelli-Ducci C, Bueno H, et al. 2017 ESC guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: the task force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). Eur Heart J. 2018;39:119-77.
6. Hellermann JP, Jacobsen SJ, Gersh BJ, Rodeheffer RJ, Reeder GS, Roger VL. Heart failure after myocardial infarction: a review. Am J Med. 2002;113:324-30.
7. Juillière Y, Cambou JP, Bataille V, Mulak G, Galinier M, Gibelin P, et al. Heart failure in acute myocardial infarction: a comparison between patients with or without heart failure criteria from the FAST-MI registry. Rev Esp Cardiol. 2012;65:326-33.
8. Libby P. Current concepts of the pathogenesis of the acute coronary syndromes. Circulation. 2001;104:365-72.
9. Wang TJ. Assessing the role of circulating, genetic, and imaging biomarkers in cardiovascular risk prediction. Circulation. 2011;123:551-65.
10. Lloyd-Jones DM. Cardiovascular risk prediction: basic concepts, current status, and future directions. Circulation. 2010;121:1768-77.
11. Cheng S, Shah SH, Corwin EJ, Fiehn O, Fitzgerald RL, Gerszten RE, et al. Potential impact and study considerations of metabolomics in cardiovascular health and disease: a scientific statement from the American Heart Association. Circ Cardiovasc Genet. 2017;10:e000032.
12. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol. 2016;17:451-9.
13. Patti GJ, Yanes O, Siuzdak G. Innovation: metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol. 2012;13:263-9.
14. Liu J, Huang L, Shi X, Gu C, Xu H, Liu S. Clinical parameters and metabolomic biomarkers that predict inhospital outcomes in patients with ST-segment elevated myocardial infarctions. Front Physiol. 2022;12:820240.
15. Oexner RR, Ahn H, Theofilatos K, Shah RA, Schmitt R, Chowienczyk P, et al. Serum metabolomics improves risk stratification for incident heart failure. Eur J Heart Fail. 2024;26:829-40.
16. Liu G, Nguyen NQH, Wong KE, Agarwal SK, Boerwinkle E, Chang PP, et al. Metabolomic association and risk prediction with heart failure in older adults. Circ Heart Fail. 2024;17:e010896.
17. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol. 2012;30:918-20.
18. Pang Z, Zhou G, Ewald J, Chang L, Hacariz O, Basu N, et al. Using MetaboAnalyst 5.0 for LC-HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. Nat Protoc. 2022;17:1735-61.
19. Shen X, Yan H, Wang C, Gao P, Johnson CH, Snyder MP. TidyMass an object-oriented reproducible analysis framework for LC-MS data. Nat Commun. 2022;13:4365.
20. Core Team RR. A language and environment for statistical computing. R Foundation for Statistical Computing. 2013
21. Ross JS, Stagliano NE, Donovan MJ, Breitbart RE, Ginsburg GS. Atherosclerosis and cancer: common molecular pathways of disease development and progression. Ann N Y Acad Sci. 2001;947:271-93.
22. Page RL 2nd, O'Bryant CL, Cheng D, Dow TJ, Ky B, Stein CM, et al. Drugs that may cause or exacerbate heart failure: a scientific statement from the american heart association. Circulation. 2016;134:e32-e69.
23. Meijers WC, de Boer RA. Common risk factors for heart failure and cancer. Cardiovasc Res. 2019;115:844-53.
24. Meijers WC, Maglione M, Bakker SJL, Oberhuber R, Kieneker LM, de Jong S, et al. Heart failure stimulates tumor growth by circulating factors. Circulation. 2018;138:678-91.
25. Kitsis RN, Riquelme JA, and Lavandero S, Heart disease and cancer: are the two killers colluding? Lippincott Williams & Wilkins Hagerstown. 2018.p.692-5.
26. Liang C, Qin Y, Zhang B, Ji S, Shi S, Xu W, et al. Energy sources identify metabolic phenotypes in pancreatic cancer. Acta Biochim Biophys Sin (Shanghai). 2016;48:969-79.
27. Moreno-Sánchez R, Rodríguez-Enríquez S, Marín-Hernández A, Saavedra E. Energy metabolism in tumor cells. FEBS J. 2007;274:1393-418.
28. Chronic Kidney Disease Prognosis Consortium; Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. Lancet. 2010;375:2073-81.

29. Smith GL, Lichtman JH, Bracken MB, Shlipak MG, Phillips CO, DiCapua P, et al. Renal impairment and outcomes in heart failure: systematic review and meta-analysis. J Am Coll Cardiol. 2006;47:1987-96.
30. Zamora E, Lupón J, de Antonio M, Vila J, Peñafiel J, Galán A, et al. Long-term prognostic value for patients with chronic heart failure of estimated glomerular filtration rate calculated with the new CKD-EPI equations containing cystatin c. Clin Chem. 2014;60:481-9.
31. Reis A, Rudnitskaya A, Chariyavilaskul P, Dhaun N, Melville V, Goddard J, et al. Top-down lipidomics of low density lipoprotein reveal altered lipid profiles in advanced chronic kidney disease. J Lipid Res. 2015;56:413-22.
32. Zhu Q, Qin M, Wang Z, Wu Y, Chen X, Liu C, et al. P Plasma metabolomics provides new insights into the relationship between metabolites and outcomes and left ventricular remodeling of coronary artery disease. Cell Biosci 2022;12:173. Erratum in: Cell Biosci. 2022;12:201
33. Qin M, Zhu Q, Lai W, Ma Q, Liu C, Chen X, et al. Insights into the prognosis of lipidomic dysregulation for death risk in patients with coronary artery disease. Clin Transl Med. 2020;10:e189.
34. Huang L, Zhang L, Li T, Liu YW, Wang Y, Liu BJ. Human plasma metabolomics implicates modified 9-cis-retinoic acid in the phenotype of left main artery lesions in acute ST-segment elevated myocardial infarction. Sci Rep. 2018;8:12958.
35. Deda O, Panteris E, Meikopoulos T, Begou O, Mouskeftara T, Karagiannidis E, et al. Correlation of serum acylcarnitines with clinical presentation and severity of coronary artery disease. Biomolecules. 2022;12:354.
36. Uddin F. Environmental hazard in textile dyeing wastewater from local textile industry. Cellulose. 2021;28:10715-39.
37. Tsukatani H, Tobiishi K. Determination of N,N-dimethyldodecylamine and N,N-dimethyloctadecylamine in river and sea water using liquid chromatography tandem mass spectrometry. Bull Environ Contam Toxicol. 2015;94:801-6.
38. Brousa E, Mastrogianni O, Orfanidis A, Zaggelidou E, Raikos N, Gika HG. Detection and determination of C12-, C14-, C16-alkyldimethylamines in human blood using gas chromatography-mass spectrometry. Rapid Commun Mass Spectrom. 2022;36:e9303.
39. Ministry of the Environment GoJ. Available from: https://www.env.go.jp/en/chemi/prtr/substances/index.html.
40. Ding J, Chen L, Shao R, Wu J, Yun Z, Dong W. Synthesis of N, N-dimethyldodecylamine by the catalytic hydrogenation of N, N-dimethyldodecylamide. React Kinet Mech Cat. 2013;108:151-9.
41. Imabeppu M, Kiyoga K, Okamura S, Shoho H, Kimura H. One-step amination of α, ω-alkylenediols over Cu/Ni-based catalysts. Catal Commun. 2009;10:753-7.
42. Mastromarino V, Casenghi M, Testa M, Gabriele E, Coluccia R, Rubattu S, et al. Polypharmacy in heart failure patients. Curr Heart Fail Rep. 2014;11:212-9.
43. Rich MW. Pharmacotherapy of heart failure in the elderly: adverse events. Heart Fail Rev. 2012;17:589-95.
44. Li Y, Yang Y, Wang X. Identification, annotation and toxicity estimation of organic pollutants in human serum via non-target analysis. Environ Pollut. 2025;367:125577.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Letter to the Editor
DOI: 10.4274/eajem.galenos.2025.88964
Eurasian J Emerg Med. 2026;25: 1

Investigation of the Efficacy of Risk Scoring Systems on Prognosis in Patients with STEMI Presenting to The Emergency Department

Erkan Boğa

Esenyurt Necmi Kadıoğlu State Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye

Keywords: ST-elevation myocardial infarction, risk scoring systems, triage in emergency departments early warning score

Dear Editor,

I read with great interest the article titled "investigation of the efficacy of risk scoring systems on prognosis in patients with ST-elevation myocardial infarction (STEMI) presenting to the emergency department," published in volume 23 of the Eurasian Journal of Emergency Medicine in 2024 (1). This study stands out for its evaluation of the effectiveness of various risk scoring systems used in emergency departments for patients with STEMI.

The findings highlight that the Triage in Emergency Departments Early Warning score (TREWS) performs as effectively as other widely used scoring systems (TIMI, ProACS, and C-ACS) in predicting short-term mortality, even demonstrating superior performance in certain instances. Particularly the high area under the curve values for TREWS (0.847 and 0.823 for 24-hour and 28-day mortality, respectively) emphasize its potential utility in clinical emergency medicine.

However, I would like to draw attention to a few points:

Study design: The single-center design and relatively small sample size may limit the generalizability of the findings. A larger, multicenter study could help validate these results more robustly (2).

Long-term outcomes: While the study focuses on short-term mortality, the evaluation of long-term outcomes (e.g., recurrent cardiac events or quality of life) would provide valuable insights for clinicians making prognostic decisions (3).

Practicality of TREWS: Additional discussion on the practical ease of using TREWS compared to other scoring systems in the time-sensitive setting of emergency departments could further highlight its clinical relevance and applicability (4).

This study provides critical insights and valuable guidance for the management of STEMI patients in emergency settings. The robust statistical analysis and methodology lend credibility to the findings. Addressing the points above in future studies may enhance the understanding and application of these risk scoring systems.

Footnotes

Conflict of Interest: No conflict of interest was declared by the author.

Financial Disclosure: The author declared that this study received no financial support.

References

1. Keski C, Yüksel M, Şensoy B, Ay MO, Avcu N, Kılıç A, et al. Investigation of the efficacy of risk scoring systems on prognosis in patients with STEMI presenting to the emergency department. Eurasian J Emerg Med. 2024;23:242-50.
2. Smith P, Johnson R, Brown T, Davis M, Miller J, Wilson A, et al. Multicenter validation of risk scores in acute coronary syndromes: implications for clinical practice. J Am Coll Cardiol. 2020;75:1456-64.
3. Anderson JL, Adams CD, Antman EM, Holmes Jr DR, Jaffe AS, Jneid H, et al. ACC/AHA guidelines for the management of patients with acute coronary syndromes: a report of the American College of Cardiology/American Heart Association Task Force. J Am Coll Cardiol. 2022;80:e223-e293.
4. Huang J, Wei X, Wang Y, Li Y, Zhang H, Liu Y, Prognostic accuracy of TIMI, GRACE, and C-ACS scores in acute coronary syndromes: a meta-analysis. Eur Heart J Acute Cardiovasc Care. 2021;10:256-64.

Corresponding Author: Erkan Boğa MD, Esenyurt Necmi Kadıoğlu State Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye
E-mail: drerkanboga@gmail.com **ORCID ID:** orcid.org/0000-0001-6802-6301

Cite this article as: Boğa E. Investigation of the efficacy of risk scoring systems on prognosis in patients with STEMI presenting to the emergency department. Eurasian J Emerg Med. 2026;25: 1.

Received: 27.01.2025
Accepted: 24.03.2025
Epub: 21.05.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.04875
Eurasian J Emerg Med. 2026;25: 8-14

Evaluation of the Relationship between Nitric Oxide Levels and Troponin, CK, CK-Mb, and CoHB Levels in Patients Presenting to the Emergency Department with Carbon Monoxide Poisoning

Bahar Keskin Çelik[1], Ali Gür[2]

[1]Manavgat State Hospital, Clinic of Emergency Medicine, Antalya, Türkiye
[2]Atatürk University Faculty of Medicine, Department of Emergency Medicine, Erzurum, Türkiye

Abstract

Aim: Ischemic alterations may occur due to hypoxia following carbon monoxide (CO) poisoning. As a result, the heart, brain, and other organs experience a loss of functionality. This study aimed to evaluate the relationship between the levels of nitric oxide (NO), which may be responsible for cardiac damage, and those of troponin, creatine kinase muscle-brain (CK-MB), and carboxyhemoglobin (COHb). These are markers of cardiac damage in patients who presented to the emergency department after being exposed to CO gas.

Materials and Methods: This prospective study included 103 individuals divided into three groups: troponin-negative patients with CO poisoning, troponin-positive patients with CO poisoning, and healthy controls. The NO levels of the groups were statistically compared. The correlation between NO levels and troponin, creatine kinase (CK), CK-MB, and COHb levels was also examined.

Results: The NO value was the lowest in the troponin-positive group and the highest in the control group, with significant differences between the three groups ($p<0.001$). According to the correlation analysis, the NO level was significantly and negatively correlated with the CK and COHb levels, but not the CK-MB level ($p<0.001$). This study revealed lower NO levels in patients with CO poisoning than in healthy individuals.

Conclusions: NO levels have an inverse relationship with troponin and CK values when comparing groups categorized by troponin levels. Therefore, it is postulated that cardiac damage can be prevented by administering inhaler NO therapy together with CO therapy to patients with CO poisoning.

Keywords: Carbon monoxide poisoning, nitric oxide, troponin, carboxyhemoglobin, cardiac damage

Introduction

Carbon monoxide (CO) poisoning is one of the most frequently reported toxicological causes of death. Poisoning occurs due to the incomplete combustion of compounds that contain carbon in their structure. After inhalation, CO passes into the blood through the lungs. When CO combines with hemoglobin, carboxyhemoglobin (COHb) is formed. The concentration of oxygen within the human body decreases as it is replaced by CO, leading to insufficient oxygen transport to tissues. As a result, the heart, brain, and other organs experience a loss of functionality (1).

Due to tissue hypoxia from CO poisoning, the release of electrons from the electron chain in mitochondrial cytochromes stops. The process of oxidative phosphorylation is compromised, leading to the occurrence of cellular hypoxia. CO also binds strongly to intracellular pigments, such as myoglobin. The toxic effects of CO on myocardial myoglobin result in a decrease in cardiac muscle contraction, thereby leading to a reduction in cardiac output. Hypoxia creates ischemic alterations in tissues, which subsequently lead to an increase in troponin, creatine kinase (CK), and creatine kinase muscle-brain (CK-MB) levels (2-3).

Corresponding Author: Ali Gür MD, Atatürk University Faculty of Medicine, Department of Emergency Medicine, Erzurum, Türkiye
E-mail: doktoraligur@gmail.com **ORCID ID:** orcid.org/0000-0002-7823-0266

Cite this article as: Keskin Çelik B, Gür A. Evaluation of the relationship between nitric oxide levels and troponin, CK, CK-Mb, and CoHB levels in patients presenting to the emergency department with carbon monoxide poisoning. Eurasian J Emerg Med. 2026;25: 8-14.

Received: 29.04.2025
Accepted: 07.06.2025
Epub: 07.07.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Nitric oxide (NO) plays an important role in regulating heart contraction, heart rate, and vascular tone. Cardiovascular diseases, such as hypertension, heart failure, ischemic heart disease, coronary artery disease, and arrhythmias, are associated with an impaired NO response (4-5). NO is of primary importance in the relaxation of coronary arteries. Furthermore, the flow-induced relaxation response in the coronary microcirculation is mediated by NO (6-7).

This study aimed to evaluate the relationship between NO levels, a potential contributor to cardiac damage, and those of troponin in patients who developed cardiac damage due to exposure to CO gas and sought medical attention in the emergency department. At the same time, an attempt was made to quantitatively reveal the relationship between CK and CK-MB values and NO values, as indicators of cardiac damage. While assessing these parameters, the correlation between the COHb values and NO values of patients with CO poisoning was also evaluated. Thus, it was investigated whether new treatment principles could be developed based on NO levels prior to the occurrence of cardiac damage in patients with CO poisoning.

Materials and Methods

Study Design and Setting

This study was conducted prospectively from September 1, 2020, to April 30, 2022, at the emergency department of a tertiary hospital (Erzurum, Health Practice and Research Hospital, Atatürk University). The study was initiated following the approval of the Ethics Committee of Atatürk Medical Faculty for Clinical Research (decision number: 53, date: 05.11.2020). Patients were accepted into the study after obtaining written informed consent from all volunteers or their relatives. Written informed consent was obtained from conscious patients, while in unconscious patients, consent was obtained in writing from their first-degree relatives. The study protocol was implemented by following the tenets of the Declaration of Helsinki.

Sample Size and Patients

The sample size of the study was calculated by using G*Power 3.1 analysis. Accepting the effect size as 0.5, the type 1 error rate as 0.05, and the power as 0.80, it was determined that a total of 90 participants (at least 30 in each group) were required. Three patient groups were formed: troponin-negative, troponin-positive, and healthy controls. A total of 30 troponin-positive and 43 troponin-negative patients presenting with CO poisoning were included in the study. Since there is no universal standard for a normal NO level, a healthy adult control group was also created, and the normal NO value was accepted as the average of the values measured in the control group. As a result, the sample consisted of 103 individuals.

Patients aged more than 18 years who were diagnosed with CO poisoning and presented to the emergency department within the first eight hours of CO exposure were included in this study. The diagnosis of CO poisoning was made based on the patients' anamnesis (medical history) and blood COHb levels above 5% (10% in smokers). The inclusion criteria of the study were as follows: being over 18 years of age, voluntary participation in the study, having a COHb level above 5 (or greater than 1 if supported by anamnesis and clinical findings), and not having a chronic disease other than hypertension. Patients excluded from the study were those aged under 18 years, those who did not volunteer to participate in the study, those with chronic diseases other than hypertension, those with other drug intoxication accompanying CO poisoning at the time of admission, pregnant women, cases in which more than eight hours had passed after CO exposure, and patients who refused treatment and follow-up. Of the total of 204 patients who presented to our emergency department with CO poisoning, 101 patients were excluded from the study because eight hours had elapsed since CO exposure, 100% oxygen therapy had been started at an external center, or due to pregnancy.

Evaluation of Patients and Data Collection

The circulation and airway scores and the Glasgow Coma Scale scores (8) of the patients were examined at the time of presentation to the emergency department, with CO poisoning.

According to this evaluation, the patients were referred to the critical care room, yellow area, or green area. Three groups were formed: the first group consisted of patients with CO poisoning who had negative troponin values (troponin-negative); the second group consisted of those with CO poisoning who had high troponin values (troponin-positive); and the third group consisted of healthy individuals without CO poisoning or chronic disease. The normal range of NO was determined based on the values measured in the control group.

The patients with CO poisoning were connected to monitors that evaluated blood pressure, respiratory rate, fever, saturation, and heart rhythm, and were placed in a security perimeter. The respiratory rate, blood pressure, pulse rate, oxygen saturation, and body temperature of the patients who were monitored in the emergency department were recorded. During this period, systemic and neurological examinations were performed. 12-lead electrocardiograms (ECGs) were taken, and the data were recorded. Levels of complete blood count, aspartate aminotransferase, alanine transaminase, arterial blood gas,

troponin, CK, CK-MB, and NO were determined as soon as possible following patient presentation. Lactate, COHb, and the presence of acidosis and alkalosis were evaluated in blood gas. The form previously prepared for patients diagnosed with CO poisoning was completed for all patients diagnosed with this condition. This form included the patient's age, gender, presentation date, emergency department protocol number, vital signs, oxygen saturation, complaints, type of CO poisoning, time since CO exposure, treatment initiation time, the presence of syncope, medical history, smoking and medication history, and arterial blood gas results. During this period, normobaric treatment (100% oxygen) was started, and hyperbaric treatment was added for those with appropriate indications. Hyperbaric treatment was planned for patients with a COHb level >25; syncope; and cardiac involvement (echocardiography findings and elevated troponin). The patients' NO levels were measured from the blood taken before hyperbaric treatment. No correlation between NO levels and hyperbaric oxygen treatment was studied. The patients were transferred to the emergency intensive care unit for follow-up.

Echocardiography was performed on the patients within the first two hours of presentation and at the 24th hour, and their ejection fraction (EF) values were measured. Cardiac changes, if any, were noted. During this process, a control ECG was taken, and any ECG changes were recorded. Echocardiography was performed by cardiologists.

Blood Collection and NO Measurement

Blood samples taken for the measurements of troponin, CK, and CK-MB levels were evaluated at the biochemistry laboratory using a Beckman Coulter AU5800 device. The data were taken from the hospital's automation system and recorded into a form. The blood samples taken for analysis of NO were transferred to 5 mL biochemistry tubes and centrifuged at 6.000 rpm for ten minutes with a Rixos 32 device. The serum obtained after centrifugation was transferred to a separate tube and stored at -80 degrees. Then, the serum was assayed collectively. A commercial enzyme-linked immunosorbent assay measurement kit (Human NO ELISA Kit, Catalog No: 201-12-1511, SunRed Biotechnology Company, China) was used to measure the NO level in the serum sample. This kit was produced for the detection of human NO levels in serum, plasma, urine, cell culture supernatant, and tissue homogenate and has a measurement range of 4-600 µmol/L, a sensitivity of 2.052 µmol/L, an intra-run precision coefficient of variation value of <10% and an inter-run precision coefficient of variation value of <12%.

Statistical Analysis

The IBM SPSS v. 20.0 (SPSS Inc., Chicago, IL, United States) program was used to analyze the data obtained. Continuous variables were expressed as mean (standard deviation) or median (interquartile range) values. Categorical data were presented as numbers (percentages). A normality analysis was performed. In the comparison between two independent groups, the independent-samples t-test was used when the normal distribution condition was met, and the Mann-Whitney U test was used otherwise. When comparing more than two independent groups and continuous variables, the ANOVA test was used for normally distributed data and the Kruskal-Wallis test for non-normally distributed data. Following ANOVA, post-hoc methods were employed using the Tukey test in cases where the variances were homogeneous and using the Tamhane T2 test in those where the variances were not homogeneous. Following Kruskal-Wallis analysis, an appropriate post-hoc test was used. However, the description of using one-way ANOVA (k-samples) may not accurately fit the context since ANOVA is not typically used as a post-hoc for Kruskal-Wallis. A p-value of <0.05 was considered significant in all statistical analyses.

Results

The study included a total of 103 patients. In the troponin-negative group, 47.7% (n=20) of the patients were female and 52.3% (n=23) were male. In the troponin-positive group, 45.2% (n=14) of the patients were female and 54.8% (n=16) were male. In the control group, 53.3% (n=16) of the patients were female and 46.7% (n=14) were male. The mean age of the entire cohort at the time of presentation was 33.44±13.82 years, and the median age was 27 (minimum: 18 - maximum: 79) years (Table 1).

Table 1 presents the patients' complaints, the presence of chest pain and syncope, the causes of CO poisoning, any ECG changes, the treatment plan, and the EF status, according to the groups.

Vital signs, pH, lactate, myoglobin values, and the EF status of the patients according to the groups are detailed in Table 2.

The comparison of the study parameters among the three groups revealed significant differences: in the COHb value between the troponin-positive group and the control group, between the troponin-negative group and the control group (p<0.001); the CK value between the troponin-positive group and the control group, between the troponin-negative group and the control group (p<0.001); the CK-MB value between the troponin-positive and troponin-negative groups (p=0.003); and the troponin value between the troponin-positive and troponin-negative groups, as well as between the troponin-negative group and the control group (p<0.001). Furthermore, the NO value significantly differed between the three groups (p< 0.001). The NO value was the lowest in the troponin-positive group and the highest in the control group (Table 3).

According to the correlation analysis, the NO level was significantly and negatively correlated with the CK and COHb levels ($p < 0.001$), but not with the CK-MB level ($p=0.265$) Furthermore, as the troponin value increased, the NO value decreased. Therefore, it can be stated that as cardiac involvement increases, the NO value decreases (Table 4).

Table 1. Demographic and clinical characteristics of the groups

		Troponin-negative (n=43, 100%)	Troponin-positive (n=30, 100%)	Control (n=30, 100%)
Age (year/mean ± SD)		32.81±11.02	39.17±18.66	28.60±9.46
Gender	Female	20 (46.5%)	14 (46.7%)	16 (53.3%)
	Male	23 (53.5%)	16 (53.3%)	14 (46.7%)
Presentation complaint	Headache	22 (51.2%)	7 (23.3%)	0 (0%)
	Dizziness	11 (25.6%)	9 (30.3%)	0 (0%)
	Fainting	1 (2.3%)	4 (13.3%)	0 (0%)
	Nausea and vomiting	8 (18.6%)	2 (6.7%)	0 (0%)
	Altered state of consciousness	1 (2.3%)	8 (26.7%)	0 (0%)
	No complaint	0 (0%)	0 (0%)	30 (100%)
Cause of CO poisoning	Stove fume	24 (55.8%)	15 (50%)	0 (0%)
	Natural gas	7 (16.3%)	7 (23.3%)	0 (0%)
	Tandoor	8 (18.6%)	1 (3.3%)	0 (0%)
	Exhaust smoke	4 (9.3%)	1 (3.3%)	0 (0%)
	Fire smoke	0 (0%)	6 (20.1%)	0 (0%)
	No known exposure	0 (0%)	0 (0%)	30 (100%)
ECG changes	Present	0 (0%)	1 (3.3%)	0 (0%)
	Not present	43 (100%)	29 (96.7%)	30 (100%)
Chest pain	Present	0 (0%)	3 (10%)	0 (0%)
	Not present	43 (100%)	27 (90%)	30 (100%)
Syncope	Present	6 (14%)	9 (30%)	0 (0%)
	Not present	37 (86%)	21 (70%)	30 (100%)
Treatment plan	Normobaric	37 (86%)	0 (0%)	0 (0%)
	Hyperbaric	6 (14%)	30 (100%)	0 (0%)
	No treatment	0 (0%)	0 (0%)	30 (100%)

SD: Standard deviation, CO: Carbon monoxide, ECG: Electrocardiogram

Table 2. Vital signs and EF status of the patients according to the groups

		Troponin-negative (n=43)	Troponin-positive (n=30)	Control (n=30)
Vital signs	GCS score (median)	15	15	15
	SBP (mmHg)	122.84±11.395	124.67±19.161	119.70±13.35
	DBP (mmHg)	69.51±8.93	67.67±9.91	66.53±9.64
	sPO_2 (%), median	97	95	98
	Pulse/min, median	80	88	80
pH, median (IQR)		7.39 (7.33)	7.38 (7.36)	7.38 (7.34)
Lactate, median (IQR)		1.4 (0.8)	3.0 (2.6)	0.950 (0.5)
Myoglobin, median (IQR)		18.4 (14.4)	31.0 (95.1)	12.5 (7.3)
EF (%), median (IQR)		60 (5)	55 (5)	60 (0)

EF: Ejection fraction, GCS: Glasgow Coma Scale, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, sPO_2: Oxygen saturation, IQR: Interquartile range

Table 3. Evaluation of the patients' NO, CK, CK-MB, troponin, and COHb levels according to the groups					
Parameter	**Group 1**	**Group 2**	**Group 3**	**p**	**Post-hoc**
NO (µmol/L)	45 (26.0)	17.5 (31.3)	110 (95)	0.000	1-2, 1-3, 2-3
Troponin (ng/LT)	2.40 (2.4)	37.0 (1875.5)	1.35 (2.7)	0.000	2-1, 2-3
CK-MB (U/L)	24 (20)	50 (101)	31 (22)	0.000	1-2
CK (U/L)	90 (75)	146 (11)	16 (9)	0.000	3-1, 3-2
COHb (%)	10.2 (15.9)	18.1 (24.3)	0.8 (0.9)	0.000	3-1, 3-2
NO: Nitric oxide, COHb: Carboxyhemoglobin, Group 1: Troponin-negative, Group 2: Troponin-positive, Group 3: Control, CK-MB: Creatine kinase muscle-brain					

Table 4. Evaluation of the relationship between NO and CK, CK-MB, COHb, and troponin in patients with CO poisoning					
	NO	COHb	Troponin	CK	CK-MB
Spearman's rank correlation	1	0-.499	0-.535	0-.537	0-.110
P		0.000	0.000	0.000	0.265
N		103	103	103	103
NO: Nitric oxide, COHb: Carboxyhemoglobin, CK-MB: Creatine kinase muscle-brain					

Table 5. Evaluation of the relationship between NO and length of hospital stay in patients with CO poisoning		
	NO	**Length of hospital stay**
Spearman's rank correlation	1	0-0.725
P		0.000
N		103
NO: Nitric oxide, Carbon monoxide (CO)		

There was an inverse correlation between the NO level and length of hospital stay in patients with CO poisoning (p=0.001). In other words, as the NO value decreased, the length of hospital stay increased (Table 5).

Discussion

In most fatalities caused by CO gas, acute myocardial damage plays an important role in increased long-term mortality. The role of NO in the regulation of heart contraction, heart rate, and vascular tone is significant. In the current study, individuals with CO poisoning who had elevated troponin levels were observed to have a significantly lower NO level compared to both troponin-negative patients and healthy controls. Furthermore, the NO value of the troponin-negative patient group was significantly lower than that of the control group. Further investigations are warranted to explore the efficacy of inhaled NO interventions as a potential measure to prevent future cardiac events and as a supplementary approach to oxygen therapy in individuals with CO poisoning, NO is considered an important intracellular and intercellular, bioactive molecule that affects various physiological and pathophysiological functions in the body, including cardiac contractility and vasodilation regulation (9). During ischemia, NO synthesis increases with increased NO synthase activity. Cardiac damage increases with cellular cytotoxicity. However, oxygenation during reperfusion facilitates delayed NO production, especially in the periinfarct region, and the conversion of NO into peroxynitrite reduces the NO level (10).

In this study, the NO level was found to be very high in the healthy adult control group, while it was the lowest in the group exhibiting elevated troponin values. The NO level also significantly differed between the troponin-negative and troponin-positive patients with CO poisoning, with lower values being detected in the latter.

In animal experiments, endothelial cells have been implicated in the development of ischemic damage. Myocardial ischemia is considered to develop after endothelial damage from CO poisoning. The disruption of the NO synthase enzyme, released from the endothelium due to endothelial damage, decreases NO synthesis. NO has an inverse correlation with ischemia-induced cardiac damage and troponin elevation (11). Similarly, in the current study, it was observed that as the NO level decreased, the troponin level of the patients with CO poisoning increased.

In a study comparing serum NO levels before and after the treatment of CO poisoning, it was found that the NO level was higher after treatment (12). This can be considered an indicator that the NO level decreases in the presence of CO exposure.

It plays an important role in regulating heart contraction, beat rate, and vascular tone. Cardiovascular diseases such as

hypertension, heart failure, ischemic heart disease, coronary artery disease and arrhythmias are associated with impaired NO response (5). Myocardial damage due to CO poisoning occurs with myocardial hypoperfusion; the CO has a toxic effect on myocardial mitochondria. Consequently, various arrhythmias, including ventricular extrasystole and fibrillation, may develop rapidly after CO exposure, without tachycardia or myocardial damage, which can affect mortality (13). In the current study, consistent with the literature, the NO level of the patients with cardiac damage was found to be lower than that of the healthy controls.

Intracellular heme protein function is affected by the partial pressure of at least three gases. These are CO, O^2 and NO. Even at low COHb levels, the CO concentration is 109 times greater than the NO concentration. CO gas binds to the heme competitively with NO. When CO binds to heme, free NO remains in the environment. This free NO has been shown to directly impair myocardial contraction (14). As the CO concentration in the environment increases, NO is expelled from heme proteins and replaced with CO (15).

CK-MB, troponin, myoglobin, and brain natriuretic peptide are the most expressed cardiac proteins in the heart muscle, and they increase after myocardial damage (16-18). The most frequently used biomarkers are troponin I, cardiovascular isoenzyme CK-MB, and CK (19-20). Icme et al. (21) reported statistically significantly higher lactate, CK-MB, and troponin I levels in patients receiving hyperbaric treatment than in those receiving normobaric treatment. In another study, Toksoy (13) detected no correlation between COHb and troponin I levels. In the same study, the mean CK level was found to be 282 U/L, and the mean CK-MB level was 27.45 U/L. In the current study, there was a negative correlation among other cardiac parameters, such as troponin, CK, and CK-MB.

The COHb level is one of the main diagnostic criteria after exposure to CO gas. CO can be produced endogenously and exogenously. In a study conducted by Toksoy (13), a statistically significant difference was found between the presentation complaints of the patients and their COHb levels. In contrast, Hampson et al. (22) found no significant correlation between presentation complaints and COHb values. Karakuzu (23), evaluating retrospective data, determined that high COHb levels did not correlate with clinical findings. Yurtseven et al. (24) detected no correlation between troponin I and COHb levels in patients with CO gas exposure and high troponin levels. In the current study, there was an inverse correlation between the NO and COHb levels, i.e., as the COHb value increased, the NO value decreased.

Study Limitations

The first limitation of our study is the very short half-life of NO, and that NO levels were not examined according to the time of presentation to the hospital in patients with CO poisoning. Moreover, several patients were transported from the incident site by ambulance, and the administration of oxygen therapy through a nasal cannula in the ambulance was a variable that influenced the NO level; however, this treatment was not included in our evaluation. Another limitation of our study is that it is a single-center study and the long-term results are unknown.

Conclusion

Patients with CO poisoning exhibit lower NO levels compared to healthy individuals. The NO level decreases further patients with high troponin levels as compared to in troponin-negative cases. Therefore, it is postulated that inhaled NO therapy in conjunction with CO therapy may serve as a preventive measure against cardiac damage in individuals exposed to CO.

Ethics

Ethics Committee Approval: The study was initiated following the approval of the Ethics Committee of Atatürk Medical Faculty for Clinical Research (decision number: 53, date: 05.11.2020).

Informed Consent: Patients were accepted into the study after obtaining written informed consent from all volunteers or their relatives.

Footnotes

Authorship Contributions

Surgical and Medical Practices: B.K.Ç., A.G., Concept: A.G., Design: B.K.Ç., A.G., Data Collection or Processing: B.K.Ç., A.G., Analysis or Interpretation: A.G., Literature Search: B.K.Ç., A.G., Writing: B.K.Ç., A.G.

Conflict of Interest: The authors declare that they have no conflicts of interest.

Financial Disclosure: This study was financially supported by the Scientific Research Projects (BAP) Unit of Ataturk University, Erzurum, TURKEY with the project code TTU-2021-8271.

References

1. Hampson NB. U.S. Mortality due to carbon monoxide poisoning, 1999-2014. Accidental and Intentional Deaths. Ann Am Thorac Soc. 2016;13:1768-74.
2. Rose JJ, Wang L, Xu Q, McTiernan CF, Shiva S, Tejero J, et al. Carbon monoxide poisoning: pathogenesis, management, and future directions of therapy. Am J Respir Crit Care Med. 2017;195:596-606.
3. Nguyen V, Salama M, Fernandez D, Sperling JD, Regina A, Rivera R, et al. Comparison between carbon monoxide poisoning from hookah smoking versus other sources. Clin Toxicol (Phila). 2020;58:1320-5.

4. Zapol WM, Rimar S, Gillis N, Marletta M, Bosken CH. Nitric oxide and the lung. Am J Respir Crit Care Med. 1994;149:1375-80.
5. Nong Z, Hoylaerts M, Van Pelt N, Collen D, Janssens S. Nitric oxide inhalation inhibits platelet aggregation and platelet-mediated pulmonary thrombosis in rats. Circ Res. 1997;81:865-9.
6. Gaston B, Drazen JM, Loscalzo J, Stamler JS. The biology of nitrogen oxides in the airways. Am J Respir Crit Care Med. 1994;149:538-51.
7. Archer SL, Huang JM, Hampl V, Nelson DP, Shultz PJ, Weir EK. Nitric oxide and cGMP cause vasorelaxation by activation of a charybdotoxin-sensitive K channel by cGMP-dependent protein kinase. Proc Natl Acad Sci U S A. 1994;91:7583-7.
8. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. Lancet. 1974;2:81-4.
9. Nager EC, O'Connor RE. Carbon monoxide poisoning from spray paint inhalation. Acad Emerg Med. 1998;5:84-6.
10. Castillo J, Rama R, Dávalos A. Nitric oxide-related brain damage in acute ischemic stroke. Stroke. 2000;31:852-7.
11. Kumar V, Cotran RS, Robbins SL. Kan damarı hastalıkları. Temel Patoloji. 1995;5:277-305.
12. Ertosun FM, Zengin S, Can B, Ulusal H, Yucel A, Taysi S. Increased nitrosative stress and oxidative DNA damage in patients with carbon monoxide poisoning. React Oxyg Species. 2020;9:136-43.
13. Toksoy A. Retrospective analysis of cardiac involvement in patients presenting with carbon monoxide poisoning to Atatürk University Faculty of Medicine Emergency Medicine Clinic. Thesis Study, Department of Emergency Medicine, Atatürk University. 2017.
14. Wunderlich C, Flögel U, Gödecke A, Heger J, Schrader J. Acute inhibition of myoglobin impairs contractility and energy state of iNOS-overexpressing hearts. Circ Res. 2003;92:1352-8.
15. Tremper KK, Barker SJ. Pulse oximetry. Anesthesiology. 1989;70:98-108.
16. Wu AHB. Release of cardiac troponin from healthy and damaged myocardium. Frontiers in Laboratory Medicine. 2017;1:144-50.
17. Gaggin HK, Januzzi JL Jr. Natriuretic peptides in heart failure and acute coronary syndrome. Clin Lab Med. 2014;34:43-58.
18. Kim JS, Kim M, Kim YJ, Ryoo SM, Sohn CH, Ahn S, et al. Troponin testing for assessing sepsis-induced myocardial dysfunction in patients with septic shock. J Clin Med. 2019;8:239.
19. Coburn RF, Ploegmakers F, Gondrie P, Abboud R. Myocardial myoglobin oxygen tension. Am J Physiol. 1973;224:870-6.
20. Katus HA, Remppis A, Neumann FJ, Scheffold T, Diederich KW, Vinar G, et al. Diagnostic efficiency of troponin T measurements in acute myocardial infarction. Circulation. 1991;83:902-12.
21. Icme F, Kozaci N, Ay MO, Avci A, Gumusay U, Yilmaz M, et al. The relationship between blood lactate, carboxy-hemoglobin and clinical status in CO poisoning. Eur Rev Med Pharmacol Sci. 2014;18:393-7.
22. Hampson NB, Dunn SL; UHMCS/CDC CO poisoning surveillance group. symptoms of carbon monoxide poisoning do not correlate with the initial carboxyhemoglobin level. Undersea Hyperb Med. 2012;39:657-65.
23. Karakuzu Engin. Investigation of carboxyhemoglobin level and clinical correlation in carbon monoxide poisoning. GATA Tıpta Uzmanlık Tezi. 2012.
24. Yurtseven S, Arslan A, Eryigit U, Gunaydin M, Tatli O, Ozsahin F, et al. Analysis of patients presenting to the emergency department with carbon monoxide intoxication. Turk J Emerg Med. 2016;15:159-62.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
DOI: 10.4274/eajem.galenos.2025.00236
Case Report
Eurasian J Emerg Med. 2026;25: 15-9

Supraventricular Tachycardia Triggered by Spicy Food Consumption: A Novel Case Report and Review of Potential Mechanisms

Kerem Sami Keyf, İlker Kaçer

Aksaray University Faculty of Medicine, Department of Emergency Medicine, Aksaray, Türkiye

Abstract

Supraventricular tachycardia (SVT) is a common cardiac arrhythmia characterized by rapid heart rates originating above the ventricles. Numerous triggers, such as caffeine, alcohol, and stimulants, have been extensively reported; however, the role of dietary spices as potential triggers remains largely unexplored. Capsaicin, the active ingredient in chili peppers, acts primarily through transient receptor potential vanilloid type 1 (TRPV1) channels, exerting diverse physiological effects, including modulation of cardiovascular function. While capsaicin has documented long-term cardiovascular protective properties, its acute arrhythmogenic potential through other mechanisms remains understudied.

We report an unprecedented case of a 48-year-old male who presented with recurrent episodes of SVT triggered consistently by spicy food intake. The patient arrived at the emergency department approximately 30 minutes after spicy food consumption with a heart rate of 215 beats per minute, resistant to vagal maneuvers. Administration of intravenous adenosine successfully restored sinus rhythm. Comprehensive cardiac evaluations, including echocardiography, coronary angiography, and electrophysiological studies, revealed no underlying cardiac pathology.

Based on this case, we hypothesize a potential role for capsaicin-induced activation of TRPV4 channels, influencing cardiac contractility and arrhythmogenesis. Prior literature has primarily addressed capsaicin's interaction via TRPV1 channels, neglecting acute electrophysiological disturbances. This novel clinical scenario underscores the need for clinicians to consider spicy food consumption as a possible trigger for SVT in differential diagnosis. Further studies are warranted to clarify acute cardiac responses and electrophysiological mechanisms triggered by capsaicin and related dietary spices.

Keywords: Supraventricular tachycardia, capsaicin, TRPV channels, spicy food, arrhythmogenesis

Introduction

Spices have been utilized for various purposes throughout human history (1). Spices play a significant role in culinary cultures worldwide, influencing numerous physiological systems, from the stimulation of chemosensory receptors through taste, aroma, and pungency, to their effects on surface membranes such as skin and mucosa, and, further impacting cardiovascular, respiratory, autonomic, and metabolic functions (2). Recent experimental and clinical studies on capsaicin, the active component of chili peppers, have demonstrated its significant role in cardiac physiology and pathologies, predominantly mediated via activation of the transient receptor potential vanilloid type 1 (TRPV1) channels (3). Piperine (from black pepper) and capsaicin have exhibited positive inotropic and chronotropic effects and demonstrated cross-tachyphylaxis in isolated rat atria (4). The known cardiac mechanism of capsaicin and the physiological role of TRPV1 channels in cardiac tissues are illustrated below (Figure 1) (5).

Supraventricular tachycardia (SVT) is defined as a tachycardia originating from above the ventricles, characterized by a heart rate exceeding 100 beats per minute. SVTs are classified based

Corresponding Author: İlker Kaçer MD, Aksaray University Faculty of Medicine, Department of Emergency Medicine, Aksaray, Türkiye
E-mail: ik0626@hotmail.com **ORCID ID:** orcid.org/0000-0002-9653-7958
Cite this article as: Keyf KS, Kaçer İ. Supraventricular tachycardia triggered by spicy food consumption: a novel case report and review of potential mechanisms. Eurasian J Emerg Med. 2026;25: 15-9.

Received: 04.04.2025
Accepted: 26.06.2025
Epub: 17.07.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

on their origin, including sinus tachycardia, atrial tachycardia, atrioventricular nodal reentrant tachycardia, and another form of reentrant tachycardia (6). SVT is usually self-limiting and generally not life-threatening; however, prolonged episodes accompanied by syncope, chest pain, or dyspnea necessitate intervention. The presence of accessory conduction pathways contributing to SVT increases morbidity and mortality risks, warranting more aggressive treatment (7). Acute interventions include vagal nerve stimulation (8), adenosine (9), calcium channel blockers (10), beta-adrenergic receptor antagonists (8), and procainamide, whereas definitive management often involves catheter ablation (11).

Although there are publications regarding cardiac rhythm disturbances triggered by spice consumption, clinicians typically do not prioritize spices when evaluating patients presenting with such disturbances. Here, we present a previously unreported case of SVT triggered by spice consumption. The informed consent form was obtained from the patient for this case report.

Case Report

A 48-year-old male presented to the emergency department (ED) with palpitations that started approximately 30 minutes after consuming spicy food and lasted for nearly one hour without relief. Upon admission, vital signs were as follows: blood pressure 110/80 mmHg, heart rate 215 bpm, and oxygen saturation 97% on room air. The patient denied any other active complaints. Medical records revealed multiple prior ED visits with similar palpitations following spicy food intake. He reported no chronic illnesses, medication use, cardiac history, or family history of sudden cardiac death. Detailed questioning revealed no alcohol, tobacco, energy drink, or stimulant use.

Physical examination showed a Glasgow Coma scale of 15, good general appearance, normal lung auscultation, tachycardic heart rhythm, without additional sounds or murmurs, and normal findings in other systems. Initial electrocardiogram (Figure 2) demonstrated SVT [hazard ratio (HR) 215 bpm]. Following unsuccessful modified Valsalva maneuvers, intravenous adenosine administration (6 mg, then 12 mg) restored normal sinus rhythm (HR 92 bpm, Figure 3). Cardiac enzymes and blood tests were within normal limits. Cardiology consultation and bedside echocardiography revealed no abnormalities. The patient remained stable without recurrence of SVT during ED observation. He was discharged after six asymptomatic hours with advice to avoid spicy foods. Subsequent elective coronary angiography and electrophysiological studies were unremarkable.

Discussion

Various underlying factors associated with SVT have been reported, with alcohol and caffeine consumption being the most common dietary triggers. However, spicy food has not been previously documented in the literature as a potential acute precipitant of SVT episodes (12). While a limited number of studies have explored the cardiovascular implications of capsaicin the active compound in chili peppers none have directly associated it with SVT (13,14).

Previous case reports and observational studies have linked spicy food consumption to a spectrum of cardiovascular and systemic effects, including syncope, urticaria, gastrointestinal symptoms, coronary vasospasm, acute myocardial infarction, and even fatal myocardial ischemia in predisposed individuals (12,15). These findings highlight that capsaicin's bioactive role extends beyond gastrointestinal discomfort and may involve acute cardiovascular modulation.

From a mechanistic perspective, capsaicin is a potent agonist of the TRPV channels, primarily TRPV1 and, to a lesser extent, TRPV4. While TRPV1 activation has been associated with long-term cardioprotective outcomes such as nitric oxide-mediated vasodilation, antihypertensive effects, and attenuation of atherosclerosis (16), its acute effects-especially in the context of electrical conduction abnormalities-remain under-investigated. Recent preclinical studies suggest that TRPV4 channels, which are also expressed in ventricular cardiomyocytes, can influence myocardial contractility and are implicated in arrhythmogenesis through calcium influx and mechanosensitive signaling pathways.

Our hypothesis builds on these findings: capsaicin may have triggered SVT in this patient via acute TRPV4 activation, leading to intracellular calcium dysregulation and enhanced automaticity or reentry mechanisms within the atrioventricular node. This is supported by limited evidence showing acute elevations in heart rate following capsaicin ingestion, although electrophysiological studies assessing arrhythmic potential are notably lacking.

Furthermore, individual susceptibility may play a significant role. Genetic variability in TRPV channel expression, underlying autonomic tone, and pre-existing subclinical conduction abnormalities might predispose certain individuals to arrhythmic events following spicy food intake. Unfortunately, the absence of baseline Holter or electrophysiological data in this patient limits causal inference.

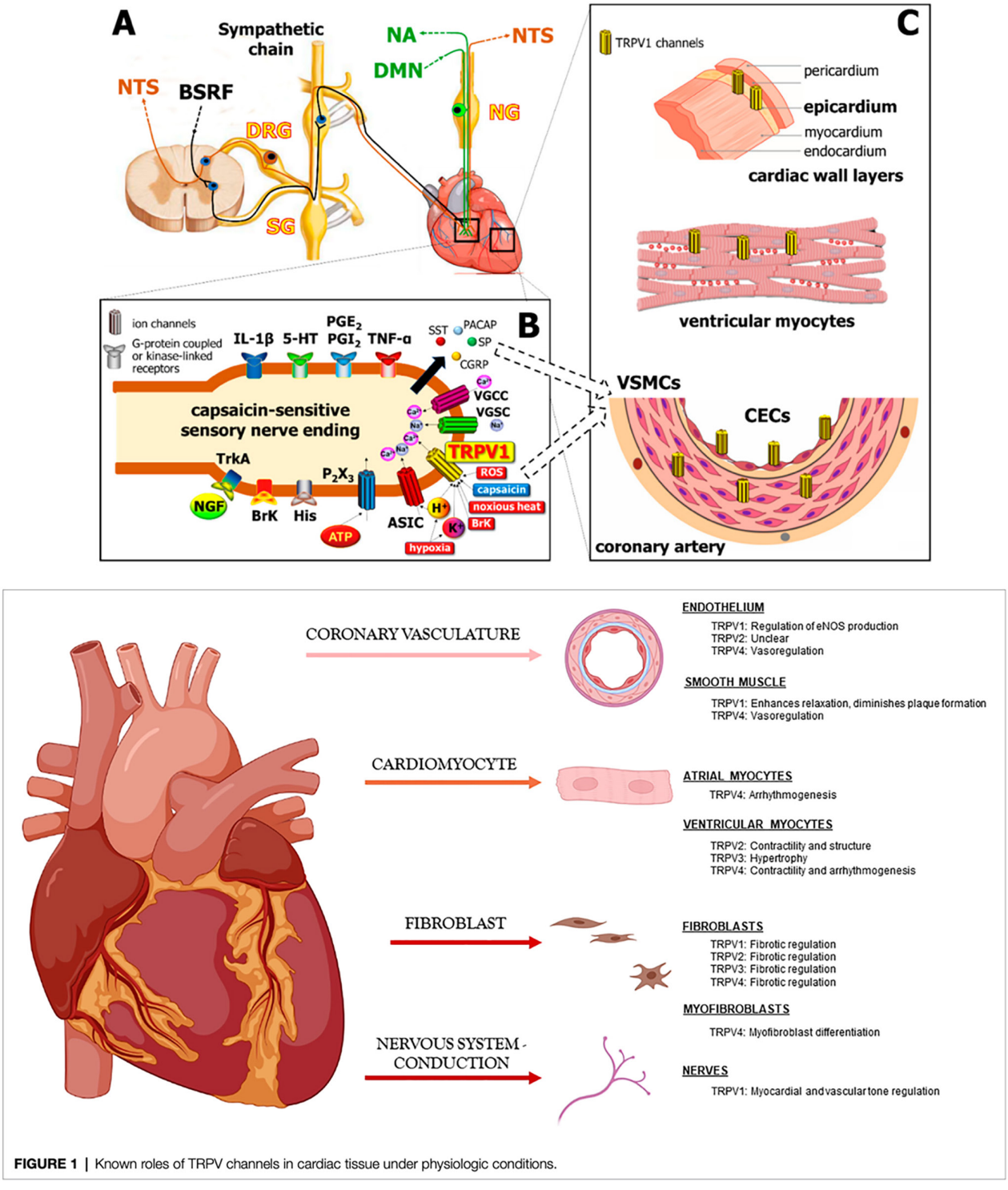


Figure 1. Known cardiac mechanism of action of capsaicin (3) and known roles of TRPV channels on cardiac tissue under physiological conditions (5)

TRPV: Transient receptor potential vanilloid, NTS: Nucleus tractus solitarius, BSRF: Brainstem respiratory field, DRG: Dorsal root ganglion, SG: Sympathetic ganglion, NA: Nucleus ambiguus, DMN: Dorsal motor nucleus, NG: Nodose ganglion, IL-1B: Interleukin-1 beta, 5-HT: 5-hydroxytryptamine, PGE: Prostaglandin E, TNF-a: Tumor necrosis factor-alpha, SST: Somatostatin, PACAP: Pituitary adenylate cyclase-activating polypeptide, SP: Substance P, CGRP: Calcitonin gene-related peptide, VGCC: Voltage-gated calcium channels, VGSC: Voltage-gated sodium channels, ROS: Reactive oxygen species, BrK: Bradykinin, ATP: Adenosine triphosphate, NGF: Nerve growth factor, VSMC: Vascular smooth muscle cell, CEC: Cerebral endothelial cell

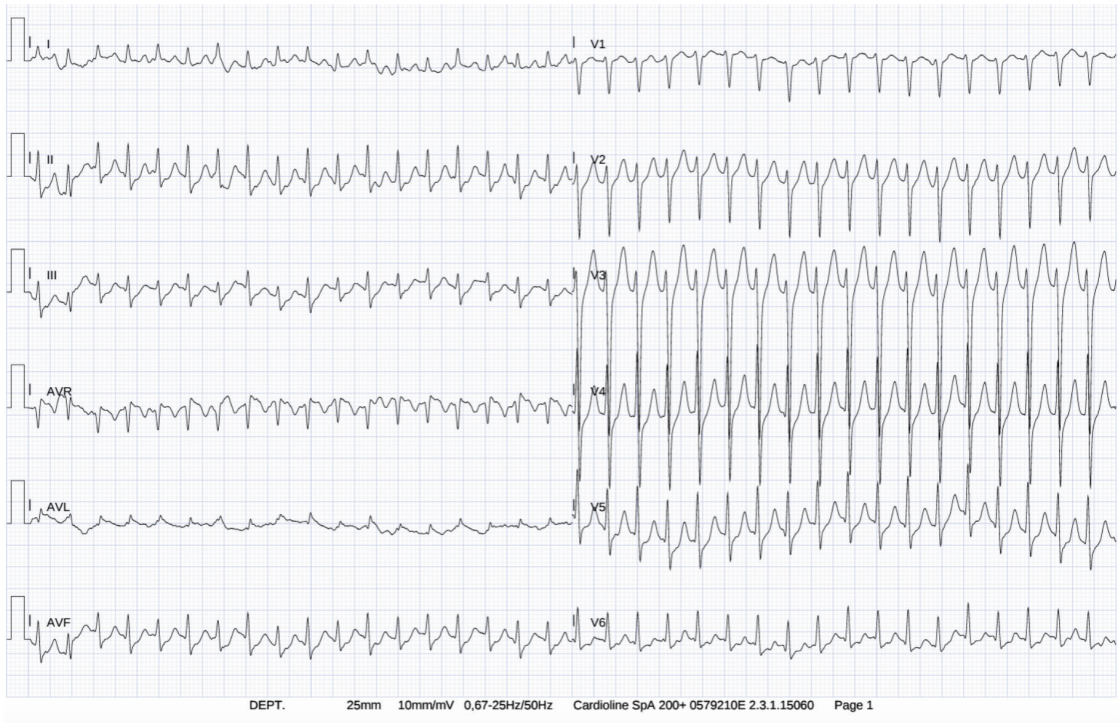


Figure 2. ECG taken at the patient's first application

ECG: Electrocardiogram, AVR: Augmented vector right, AVL: Augmented vector left, AVF: Augmented vector foot

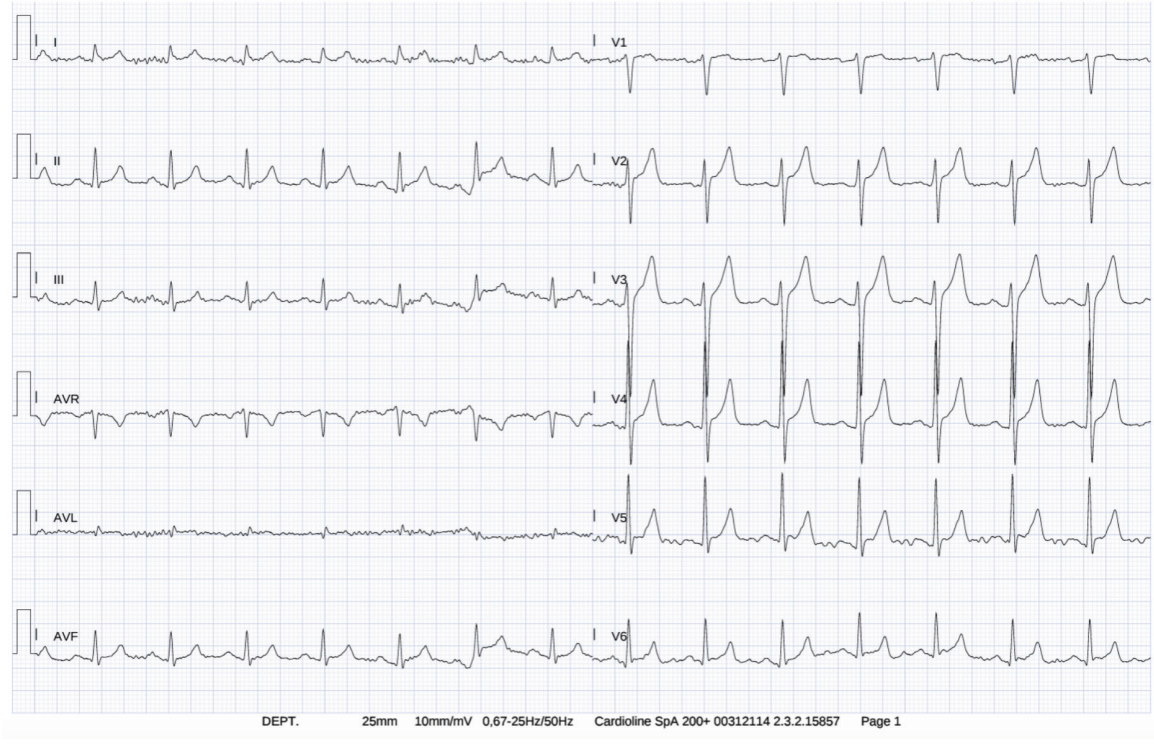


Figure 3. Control ECG taken after 6-12 mg intravenous adenosine treatment

ECG: Electrocardiogram, AVR: Augmented vector right, AVL: Augmented vector left, AVF: Augmented vector foot

Conclusion

In conclusion, this case introduces a novel and plausible trigger for SVT that has been previously overlooked in emergency clinical settings. It underscores the need for a more comprehensive dietary history in patients presenting with idiopathic arrhythmias. Future prospective studies are warranted to investigate the acute electrophysiological effects of capsaicin and delineate the specific roles of TRPV1 and TRPV4 channels in cardiac rhythm regulation.

Ethics

Informed Consent: The informed consent form was obtained from the patient for this case report.

Footnotes

Authorship Contributions

Concept: İ.K., Data Collection or Processing: K.S.K., Writing: İ.K., K.S.K.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Jack Turner. Spice: The history of a temptation. 2005.
2. Kochar K. Effects of dietary herbs and spices. Journal of Orthomolecular Medicine. 1999;14:210-8.
3. Szabados T, Gömöri K, Pálvölgyi L, Görbe A, Baczkó I, Helyes Z, et al. Capsaicin-sensitive sensory nerves and the TRPV1 Ion channel in cardiac physiology and pathologies. Int J Mol Sci. 2020;21:4472.
4. Miyauchi T, Ishikawa T, Sugishita Y, Saito A, Goto K. Involvement of calcitonin gene-related peptide in the positive chronotropic and inotropic effects of piperine and development of cross-tachyphylaxis between piperine and capsaicin in the isolated rat atria. J Pharmacol Exp Ther. 1989;248:816-24.
5. Peng G, Zei PC. Diagnosis and management of paroxysmal supraventricular tachycardia. JAMA. 2024;33:601-10.
6. Waldo AL. Atrial fibrillation and atrial flutter: two sides of the same coin! Int J Cardiol. 2017;240:251-2.
7. Whinnett ZI, Sohaib SM, Davies DW. Diagnosis and management of supraventricular tachycardia. BMJ. 2012;345:e7769.
8. Chauhan VS, Krahn AD, Klein GJ, Skanes AC, Yee R. Supraventricular tachycardia. Med Clin North Am. 2001;85:193-223.
9. Fox DJ, Tischenko A, Krahn AD, Skanes AC, Gula LJ, Yee RK, et al. Supraventricular tachycardia: diagnosis and management. Mayo Clin Proc. 2008;83:1400-11.
10. Supraventricular tachycardia: diagnosis and management. Mayo Clin Proc. 2008;83:1400-11.
11. Soejima K, Delacretaz E, Suzuki M, Brunckhorst CB, Maisel WH, Friedman PL, et al. Saline-cooled versus standard radiofrequency catheter ablation for infarct-related ventricular tachycardias. Circulation. 2001;103:1858-62.
12. Li Q, Chang M, Lai R, Zhang H, Song L, Wang X, et al. Potential benefits of spicy food consumption on cardiovascular outcomes in patients with diabetes: a cohort study of the China Kadoorie Biobank. Nutrition. 2023;112:112062.
13. Ofori-Asenso R, Mohsenpour MA, Nouri M, Faghih S, Liew D, Mazidi M. Association of spicy chilli food consumption with cardiovascular and all-cause mortality: a meta-analysis of prospective cohort studies. Angiology. 2021;72:625-32.
14. Li J, Xie C, Lan J, Tan J, Tan X, Chen N, et al. Spicy food consumption reduces the risk of ischaemic stroke: a prospective study. Br J Nutr. 2024;131:1777-85.
15. Lei J, Zhu F, Zhang Y, Duan L, Lei H, Huang W. Transient receptor potential vanilloid subtype 1 inhibits inflammation and apoptosis via the release of calcitonin gene-related peptide in the heart after myocardial infarction. Cardiology. 2016;134:436-43.
16. Miller M, Koch SE, Veteto A, Domeier T, Rubinstein J. Role of known transient receptor potential vanilloid channels in modulating cardiac mechanobiology. Front Physiol. 2021;12:734113.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
DOI: 10.4274/eajem.galenos.2025.50024
Review
Eurasian J Emerg Med. 2026;25: 20-8

Practical and Legal Aspects of Applying Artificial Intelligence in Emergency Medicine

Tomasz Buczek[1], Ilon Szwedziak[2], Emin Fatih Vişneci[3], Michal Pruc[4], Francois-Xavier Duchateau[5], Anne Lepetit[5], Lukasz Szarpak[6]

[1]Victorialex Law Firm J. Piejko Limited Partnership, Warsaw, Poland
[2]University of Warsaw Faculty of Management, Warsaw, Poland
[3]Konya City Hospital, University of Health Science, Department of Emergency Department, Konya, Türkiye
[4]LUXMED Group, Clinic of Clinical Research and Development, Warsaw, Poland
[5]BUPA Group, London, UK
[6]The John Paul II Catholic University of Lublin, Institute of Medical Science, Collegium Medicum, Lublin, Poland

Abstract

This narrative review explores current applications of artificial intelligence (AI) in emergency medicine, critically evaluates the supporting evidence, and discusses the ethical, legal, and regulatory challenges surrounding its integration into clinical practice. Peer-reviewed literature and recent systematic reviews on AI applications in emergency medicine were analyzed using a structured narrative approach. AI-driven operational forecasting, predictive modeling for patient outcomes, diagnostic support, and AI-assisted triage systems are among the domains evaluated. AI models, such as neural networks and gradient boosting machines, have demonstrated superiority over traditional triage tools in forecasting outcomes like in-hospital mortality and intensive care unit admission. AI in diagnostics has enhanced point-of-care ultrasound analysis, sepsis detection, and electrocardiogram interpretation. Operationally, AI makes it possible to predict patient volume, emergency department crowding, and resource allocation in real time. Despite these developments, there are still few prospective clinical trials confirming better patient outcomes. Algorithmic bias, a lack of transparency, automation bias, and restrictions on generalizability across clinical settings are among the main issues. Emerging regulatory frameworks like the European Union AI act and ethical and legal frameworks like the General Data Protection Regulation and Health Insurance Portability and Accountability Act are essential for directing the responsible use of AI. AI has significant potential to improve the provision of emergency care. However, ethical protections, legal compliance, integration with clinical workflows, and thorough external validation are necessary for responsible implementation. To guarantee the safe and fair implementation of AI in emergency medicine, future initiatives must concentrate on explainable AI, multicenter prospective research, and stakeholder collaboration.

Keywords: Artificial intelligence, emergency medicine, clinical decision support, triage systems, machine learning, ethical issues, legal regulation, prognostic models, diagnostic accuracy

Introduction

Emergency medicine is a high-stakes profession with inherent resource limitations, complicated patient flow, and time-sensitive decision-making. Chronic emergency department (ED) overcrowding and the Coronavirus disease-2019 pandemic have brought attention to systemic issues like long wait times, employee burnout, and poor patient outcomes (1). Artificial intelligence (AI) has become a promising tool to supplement emergency care in this context. AI systems can help with quick triage, diagnostic support, outcome prediction, and operational decision-making thanks to developments in machine learning (ML) and data analytics. In fact, AI methods have already shown potential in enhancing

Corresponding Author: Lukasz Szarpak Prof. The John Paul II Catholic University of Lublin, Institute of Medical Science, Collegium Medicum, Lublin, Poland
E-mail: lukasz.szarpak@gmail.com **ORCID ID:** orcid.org/0000-0002-0973-5455

Cite this article as: Buczek T, Szwedziak I, Vişneci EF, Pruc M, Xavier Duchateau F, Lepetit A. Practical and legal aspects of applying artificial intelligence in emergency medicine. Eurasian J Emerg Med. 2026;25: 20-8.

Received: 17.06.2025
Accepted: 25.07.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

clinical decision support, medical imaging interpretation, and ED triage accuracy (1,2). From prehospital evaluation to in-hospital diagnostics, prognostication, and resource management, AI technologies are being used increasingly throughout the emergency care spectrum. AI-enabled symptom checkers can be used by patients before they arrive, and dispatch centers can use AI to streamline response logistics and triage. AI models assist in clinical decision-making during hospitalization by predicting results, analyzing diagnostics, and classifying risk. Predictive algorithms are used administratively to guide strategies for crowding mitigation, bed distribution, and staffing.

It is still difficult to integrate AI advancements into standard emergency care, despite this promise. Relatively few prospective trials have shown better patient outcomes in real-world settings, and the majority of AI applications in emergency medicine to date have come from retrospective studies or proof-of-concept models (1,2). For example, a 2020 scoping review identified 150 studies of AI in emergency care, over 82% of which were retrospective and only 2% were prospective controlled trials (2). The overall body of evidence is still small, even though about 25% of these interventions sought to improve diagnosis (particularly in imaging), and some algorithms even beat physicians in particular tasks. Before AI tools are widely used, thorough validation and an evaluation of their effects on patient-oriented outcomes are still required. This narrative review explores the main ethical, legal, and regulatory concerns related to the use of AI in emergency medicine while acknowledging both the potential and the gaps. The methodological approach, major application domains (triage, diagnostics, prognostication, resource management), ethical considerations (fairness, transparency, oversight, consent), and the legal/regulatory landscape [liability, data protection, and governance frameworks in the United States (US) and European Union (EU)] are all covered.

Triage

An important advancement in digital health is the growing availability of AI-based symptom checkers on web and mobile platforms, which provide users with real-time triage recommendations (3). These tools interpret user-provided symptoms by utilizing natural language processing (NLP) and pattern recognition algorithms recommend appropriate care-seeking actions based on perceived urgency (4). The scientific community continues to have concerns about their accuracy and safety, despite their promising potential to empower patients and manage healthcare demand (3).

Variability in these systems' performance has been brought to light by systematic reviews and comparative studies. For example, a 2020 study that looked at 22 different symptom checkers found significant variations in triage recommendations and diagnostic accuracy (3). Interestingly, the study found that both under- and over-triage rates were high, especially for conditions that require immediate attention, like myocardial infarction or stroke (3). While over-triage can unnecessarily strain emergency services, under-triage can result in delayed care for critical conditions, potentially worsening outcomes (5).

The proprietary nature of many underlying algorithms, sometimes referred to as "black boxes", presents a significant challenge in assessing and enhancing these tools (3). These algorithms might use antiquated medical guidelines or heuristics that haven't been thoroughly tested (3). It is challenging to evaluate their internal reasoning and spot any biases or limitations because of this lack of transparency.

Experts emphasize the importance of extensive validation studies and external benchmarking to ensure clinical reliability before public health initiatives or emergency medical care triage systems are widely implemented (3). Recalibration and prospective testing using patient data are ideal for this validation to preserve accuracy over time. Using ML and NLP to improve emergency room triage precision and consistency is still being studied (5). Frameworks are also being developed to reflect real-world patient circumstances in symptom assessment tool case vignettes (6). AI is being studied for clinical decision support systems, patient triage, and diagnostic support in emergency care (7). ML algorithms are also being used to predict inpatient admissions using ED triage data to improve patient flow management (8). Development of interpretable ML models for triage that address class imbalance is also underway to improve accuracy and clinical judgment (9).

More advanced AI applications are appearing in emergency dispatch. The Corti platform helps emergency call operators identify cardiac arrest-related clinical phrases, tone, and metadata using ML and real-time speech recognition. In retrospective tests, such technologies detected out-of-hospital cardiac arrest faster than human dispatchers, enabling advanced life support (10). AI simplifies triage decision-making and automatically extracts demographics and geography. Few prospective trials exist, and further study is needed to address false alarm rates, dispatcher over-reliance, and system latency before widespread adoption.

Studying the specific developments, difficulties, and factors surrounding AI-assisted triage systems is important to build on the analysis of the traditional Emergency Severity index (ESI) and the potential of AI models (like gradient boosting machines and neural networks) to improve triage accuracy.

Gradient boosting machines and neural networks can predict important outcomes such as intensive care unit (ICU) admission and in-hospital mortality better than the ESI. Studies have shown that ML models outperform the ESI in AUROC scores (11). One study using data from over 189,000 emergency patients created and validated interpretable ML models for ICU admission by comparing models based on ESI, vital signs, and a mix of vital signs, demographics, and medical history (12). Another study examined feed-forward neural networks, regularized regression, random forests, and gradient-boosted trees for predicting ICU versus non-ICU care after 24 hours of admission using 41,654 ED visits (12). These findings show that AI models can manage complex interactions between several variables, which could lead to more nuanced risk categorization than rule-based systems. The effectiveness of gradient boosting machines is demonstrated by increased research into specific ML methodologies. A study that used a gradient boosting ML model to predict early mortality in ED triage showed how useful it is of the model for enhancing patient classification (13). Extreme-Gradient-Boosting-based interpretable ML models have been used to forecast extended wait times in the ED, enabling the assessment of equity among various patient groups (14). For clinical adoption, this emphasis on interpretability is essential because it helps medical professionals comprehend the logic behind an AI's recommendation.

Even with little information available, deep-learning techniques, particularly neural networks, have demonstrated promise in identifying critically ill patients during triage (15). In time-sensitive triage situations where only preliminary information is available, deep learning's capacity to automatically extract complex features from raw data may be especially helpful. The efficiency of various ML models, such as artificial neural networks, for predicting ESI levels has also been compared (16).

The generalizability of AI models is still a major obstacle, despite the encouraging outcomes. Numerous models are trained and validated using datasets from individual institutions, which might not account for the differences in patient demographics, clinical procedures, and data gathering techniques among various emergency rooms (11). Before deployment, thorough local validation is required due to this limitation. A crucial area of study is the significance of external validation and the possibility of performance deterioration in unfamiliar settings.

Ensuring demographic fairness in AI-assisted triage is another significant challenge. Models that consistently under- or over-triage particular demographic groups, like minority populations, may result from bias in training data (14). This issue is brought to light by studies that examine differences in ED prioritization according to demographic traits even when triage acuity scores are comparable (17). When developing, assessing, and implementing models, demographic fairness must be carefully taken into account. Building equitable AI systems requires methods for assessing fairness, like those employed in the study looking at ethnic differences in wait time prediction models (14).

A number of tactics are suggested to lessen these difficulties and make it easier for AI to be used responsibly in triage. To guarantee performance and dependability, local validation of AI models within the particular context of the implementing institution is crucial (14). Incorporating human-in-the-loop supervision is also crucial. Clinical judgment should not be replaced by AI; rather, it should be seen as an aid. When needed, medical professionals should be able to override AI recommendations and offer input, to help the system improve over time.

Beyond demographic data and vital signs, multimodal data integration offers additional promise for improving AI triage accuracy. A more complete picture of a patient's condition can be obtained by using NLP to incorporate unstructured data from electronic health records (EHR), such as doctor notes and past medical information (18). In addition to data gathered at the triage point, studies have investigated the use of ML predict hospital admission at triage based on patient history (19).

Building interpretable AI models is also essential to encouraging adoption and trust among medical professionals. Clinicians can validate a recommendation based on their medical expertise when they understand why a model makes a specific prediction. There is still research being done on interpretable ML models for triage prediction (12).

Diagnostics

Advances in medical imaging, data synthesis, deep learning, and specialist tools like electrocardiogram (ECG) and sepsis tools affect pre-hospital and ED emergency diagnosis. Emergency medicine's high stakes and time limits demand fast and accurate diagnosis to improve patient outcomes (20). ED triage and outcome prediction using ML and AI are major areas of focus. ML models are being created to predict hospital admission and mortality using triage data as such as patient history, vital signs, and nursing notes (21). To reduce subjectivity in conventional systems like the ESI, a deep learning approach that uses electronic medical records (EMRs) has been proposed to improve ED triage accuracy (22). Comparisons of ML models with classical ESI show promising clinical outcome prediction (11). NLP is essential for extracting meaningful information from unstructured text data in triage notes and improving prediction accuracy (23).

Emergency diagnosis requires medical imaging. In hemodynamically unstable patients who may not be candidates

for computed tomography pulmonary angiography, point-of-care ultrasound is increasingly employed to diagnose large pulmonary embolism (24). AI models are also being studied for real-time ultrasound picture interpretation in emergency settings like the extended focused assessment with sonography in trauma test to improve diagnostic precision and decision-making (25). Multimodal AI systems that use metadata and ocular images, for primary diagnosis and eye emergency triage, show the potential of combining data sources for diagnostic efficiency (25).

Emergency cardiac diagnosis requires ECG analysis. AI and deep learning models, notably convolutional neural networks (CNNs), are revolutionizing ECG analysis by automating high-precision arrhythmia identification (25). ML algorithms using 12-lead ECGs are being developed to predict acute mortality in ED patients (25). Acute coronary syndrome patients need quick ECG acquisition at triage, and clinical prediction methods in tablet apps are being studied to accelerate this process (26). Bedside assays for high-sensitivity troponin I are also being compared to central laboratory analysis to determine if they can quickly diagnose acute myocardial infarction (27). Sepsis is life-threatening, and delayed treatment increases mortality; therefore, ED triage must discover it early (28). ML improves sepsis triage detection before lab results (28).

Data synthesis, which integrates multimodal data from physiological signals, EMRs, and medical imaging, is boosting diagnostic accuracy (29). Hybrid deep learning architectures using CNNs, recurrent neural networks, and transformer models are being studied for multimodal data fusion in healthcare diagnostics (29). To reduce ED overcrowding and improve resource allocation, deep learning is being used to analyze heterogeneous medical data to predict patient criticality and identify the appropriate clinical departments (30).

Despite advances, emergency AI-driven diagnostic tools are still challenging to implement. These include addressing the medicolegal implications of deploying AI in high-risk scenarios, ensuring AI model interpretability and fairness, and undertaking clinical validation studies (31). To ensure successful incorporation into healthcare workflows, clinician views on AI in emergency triage must be considered (31). Multidisciplinary panels must be used to investigate diagnostic errors in EDs to improve patient safety (32).

Prognostication

Emergency medicine uses AI for prognostication, or patient outcome prediction from structured data. In high-acuity, fast-paced settings like the ED, AI-driven predictive models can help doctors identify patients at risk of deterioration, ICU admission, protracted hospitalization, or death. These techniques improve early risk categorization for faster interventions, triage, and resource allocation.

ML algorithms can predict clinical worsening using preliminary ED data including vital signs, lab results, and patient demographics. Algorithms can predict in-hospital cardiac arrest, mechanical ventilation, major trauma transfusions, and infection, or septic shock. AI techniques for early sepsis identification that combine lab measurements and dynamic vital sign trajectories to provide real-time risk scores have garnered attention. Some models outperform rule-based systems in retrospective validations with AUROC values close to 0.90 (33). A meta-analysis of prognostic AI in emergency settings found that ML models outperformed conventional risk stratification methods in predicting hospital admission and short-term death (34). Many of these models are "early warning systems" that detect clinically stable patients who could decompensate in hours. Several health organizations are using AI-based early warning capabilities in their ED information systems to flag high-risk patients for clinical reassessment or escalation.

Translation into clinical practice is problematic. The generalizability of predictive AI models is often limited by their training setting. One often reported proprietary sepsis prediction system utilized in numerous hospitals performed worse than earlier reports, exhibiting low sensitivity and inappropriate calibration in external validation cohorts (35). This mismatch emphasizes the need for external validation and post-deployment monitoring. Audits, recalibrations, and impact analyses should be done regularly to detect performance drift and maintain clinical relevance.

Another issue is automation bias—physicians may overuse algorithmic risk rankings instead of clinical judgment. AI results should be considered as supporting data, not as final instructions. The doctor must interpret prognostic AI by combining contextual knowledge of the patient's presentation with the algorithm's detection of subtle, data-driven risk patterns.

Despite these shortcomings, prognostic AI could transform emergency treatment. Due to their ability to assess high-dimensional, temporally linked data, models can detect early signs of deterioration as such as hemodynamic microtrends or subtle laboratory abnormalities that humans cannot. These insights can reduce unnecessary outcomes, personalize monitoring intensity, and improve patient disposition decisions (ED discharge vs. observation vs. ICU stay). Future efforts should

focus on explainable AI (XAI) frameworks that integrate easily into ED procedures and provide interpretable risk estimates. Research priorities include prospective, multicenter trials to determine how AI-informed prognostication influences clinician judgment, patient outcomes, and health system performance.

AI is increasingly employed to improve clinical decision-making and operational efficiency in EDs, which often have congestion, resource constraints, and unpredictable patient surges. Dynamic capacity planning using AI-based forecasting, scheduling, and logistics systems may improve patient flow, reduce waits, and employee burnout for emergency medicine administrators.

Predictive Forecasting and Capacity Planning

AI demand forecasting is among its most advanced operational uses. ML models trained on historical ED data, including chief complaints, patient arrival times, seasonal trends, weather data, and public event calendars, can effectively anticipate hourly or daily patient loads. These short-term estimates enable proactive personnel, space, and bed modifications. If an AI tool predicts a high-volume day after a major athletic or meteorological event, clinical directors can plan extra staff or pre-open surge capacity units. A recent study shows that AI-based ED crowding forecasts improve throughput and reduce left without being seen (36). Heuristic-based models do not.

Triage vitals, presenting complaint, first labs, and imaging can also be used by AI to predict patient-level length of stay (LOS). This aids downstream bed planning, transfer prioritization, and real-time patient streaming. In high-occupancy settings, LOS prediction helps estimate discharge rates more precisely, which helps manage bed turnover (37).

Staff Scheduling and Workload Balancing

AI can help schedule staff by recognizing demand trends and matching staffing to expected acuity and volume. Dynamic scheduling algorithms can identify understaffed periods, optimize shift timing, and ensure the right skill mix of attending physicians, residents, and nurses during peak hours. In one application, AI-driven dashboards predicted ED crowding eight hours in advance and recommended real-time staffing adjustments, enhancing throughput and employee satisfaction (38).

AI is being studied for real-time ambulance dispatch optimization. AI algorithms can use prior EMS data, traffic patterns, and projected ED congestion to route ambulances to facilities with the correct acuity and capacity. This could reduce ED offload and transport delays, which cause system bottlenecks (39).

Inventory and Supply Chain Optimization

AI helps with supply forecasting and logistics in emergency care, in addition to managing staff and beds. Based on historical and current data, algorithms can forecast trends in resource consumption, such as the use of tPA in stroke or the need for transfusions in trauma. By using these insights, hospitals can better stock equipment, drugs, and blood products, minimizing shortages during surges and cutting waste during lulls. This type of predictive logistics is particularly useful in high-stakes situations like disaster response or mass casualty incidents (40).

Implementation Considerations

Even with promising pilot results, a number of crucial elements must come together for AI to be successfully implemented in ED operations:

Data integration: In order to access real-time data feeds, AI systems must easily interface with current hospital command centers, admission-discharge-transfer systems, and EHR.

Interpretability: The results of operational AI must be clear and intelligible. To trust and follow predictions and recommendations, clinicians and administrators must comprehend the reasoning behind these predictions and recommendations.

Stakeholder buy-in: Frontline employees' and administrative leadership's participation is crucial. To prevent resistance or unintentional disruption, tools must be co-designed with users in mind and customized to local workflows.

Small but significant gains in ED LOS, patient satisfaction, and clinician workload distribution have been shown in early trials of AI-supported operational systems (41). Scalable and validated AI tools for capacity management are set to become just as important as clinical algorithms in promoting patient-centered, resilient emergency care, as the demand on emergency services continues to increase.

Ethical Considerations in the Use of AI in Emergency Medicine

AI in emergency medicine could boost efficiency and outcomes, but ethical issues must be addressed to ensure patient safety. AI ethics must be built on fairness, openness, human monitoring, and informed consent. This is because EDs treat vulnerable patients under time and budget constraints.

Fairness and bias are fundamental ethical dilemmas. If trained on unrepresentative datasets or structural inequities, AI models may unintentionally exacerbate care disparities. If training data show that women have historically underreported cardiac symptoms, a triage algorithm may continually underestimate

risk in female patients. In rural or resource-constrained contexts, algorithms based on metropolitan university hospital data may perform poorly. Variable measurement, algorithm design, and data sampling biases might worsen healthcare inequities and discriminatory treatment. To solve this problem, datasets must be representative and diversified, and models must use fairness requirements or employ bias correction methods, such as resampling or reweighting, to ensure fairness. After deployment, institutions should evaluate demographic subgroup AI performance and establish redress processes for inequalities. Finally, equitable results for all groups served are as crucial as technical performance. Transparency matters too. Many powerful AI models, especially deep learning-based ones, are "black boxes" that predict without explanation. Emergency clinicians must understand, contextualize, and question AI-generated recommendations to make rapid choices based on limited data. XAI provides a partial solution by producing interpretable outputs such as feature attribution scores or visual explanations like diagnostic image heatmaps. These arguments improve clinician trust, accountability, and decision-making. The EU General Data Protection Regulation (GDPR) and the impending AI act require individuals to have access to relevant information about automated judgments, making explainability in high-risk AI systems increasingly critical. Clarity and usability must be balanced in practice; too comprehensive an explanation can overwhelm doctors, while an explanation that is too simple will not impart meaningful information. Understanding that hypotension and altered mental status triggered a triage signal can help make care decisions in the high-pressure ED without delaying action.

Clinician autonomy and oversight are also ethical requirements. AI should complement human decision-making, not replace it. Doctors may overuse AI results and trust AI advice without adequate critical thinking, raising concerns about automated bias. This can be troublesome if the model is used outside its intended scope or if AI recommendations clash with clinical intuition. Over time, uncritical AI use can cause "moral deskilling," leading to medical practitioners losing confidence in their judgment. AI systems reduce this risk by incorporating human-in-the-loop protections that let doctors change or question outcomes. Clinicians should have final-say over high-stakes care decisions per institutional policy. Certain AI platforms require clinicians to disclose their reasoning for rejecting or approving algorithmic advice to encourage reflection and discourage mindless acceptance. Regulators are codifying these expectations as the EU AI act requires human oversight for high-risk medical AI systems. Emergency medicine regulations may require human providers to verify AI triage scores or diagnostic signals before acting.

AI complicates informed consent, which has long been a cornerstone of moral medical practice. AI in treatment should be explained to patients, especially if it affects their decision-making. AI tool discussions in the ED are often infeasible due to time constraints, unconscious patients, and possibly lethal scenarios. These instances fall under emergency informed consent exceptions in US and European law. We should be transparent about AI's contribution wherever possible. Hospitals may employ signage, general consent forms, or institutional regulations to disclose AI systems for clinical decision-making. Hospitals should notify patients after AI significantly affects a diagnosis or treatment plan, especially without human review. Hospitals might adopt a strategy of "ogoing informed transparency" to ensure that patients are aware of AI use, even if particular disclosures are not possible during an urgent visit. Transparency is crucial when AI programs are experimental or make judgments without human input, such as EMS routing. Formal study methods or separate permission may be ethical in certain cases, although clarity on the context of 'separate permission' is advised.

Legal and Regulatory Issues

A thorough grasp of current frameworks and new regulations is necessary to navigate the substantial legal and regulatory challenges associated with the integration of AI into emergency medicine. Although AI has the potential to improve productivity and patient care, its application is limited by a complicated framework that includes liability, data protection, and regulatory supervision (42).

Laws pertaining to privacy and data protection are a major concern (43,44). The Health Insurance Portability and Accountability Act (HIPAA) in the US establishes guidelines for safeguarding private patient health data (44). Regarding the use and disclosure of protected health information, HIPAA places duties on covered entities (healthcare clearinghouses, health plans, and providers) and their business associates. HIPAA compliance is crucial when AI systems handle patient data, necessitating strong administrative and technical safeguards to guarantee data security and privacy (45).

Even more stringent regulations are enforced by the GDPR of the EU (46) which gives people broad control over their personal data, including the ability to access, amend, and remove data as well as the ability to object to processing (47). Strict adherence to the principles of data minimization, purpose limitation, and transparency in algorithmic processing is required for the

GDPRs application to AI in healthcare (46). Obtaining appropriate patient consent for data usage, establishing clear data governance policies, and ensuring adequate anonymization or pseudonymization of data used for AI training and deployment are among the legal challenges in this area (48).

Another crucial legal issue is liability, and another is accountability (49,50). The advent of AI complicates traditional medical malpractice law, which holds healthcare organizations and individual practitioners accountable for careless acts (50). Determining who is at fault—the doctor, the healthcare facility, or the AI developer—when an AI system causes a negative outcome is a complex legal matter (49). The "learned intermediary" doctrine, which holds that AI is a decision-support tool and that the doctor is ultimately in charge of assessing AI recommendations and reaching clinical conclusions, is frequently invoked in current legal viewpoints (50). Physicians are required by this doctrine to use their own clinical judgment and evaluate AI systems' output critically (50).

However, product liability law becomes more important as AI systems in emergency medicine become more self-sufficient and require less human intervention (51). The AI developer or manufacturer may be held accountable if damage is caused by an error in the warnings, design, or manufacturing of the AI system. This puts the AI product itself front and center instead of the doctor's actions. There are particular legal difficulties in establishing a flaw in a sophisticated, self-learning AI system, such as establishing causation and pinpointing the precise part or algorithm that caused the damage (49). Globally, the legal frameworks pertaining to liability protection and AI service businesses are changing, with various jurisdictions investigating potential solutions (52).

The US Food and Drug Administration (FDA) and the US, through the proposed AI act, are playing important roles in the ongoing development of regulatory pathways for AI in healthcare. In recognition of the need for frameworks that can adjust to the iterative nature of AI development and learning, the FDA has started to develop regulatory approaches for medical devices that use AI. High-risk AI applications in healthcare are subject to strict requirements for conformance assessment, risk management, data governance, and human oversight (53,54). The EU AI act, a comprehensive regulatory framework for AI, classifies AI systems according to their risk level. The goal of these regulatory initiatives is to guarantee the ethical, safe, and efficient use of AI in clinical settings; however, developers and healthcare providers face compliance challenges as a result of these changing regulations (50).

Additional legal factors include the possibility of algorithmic bias producing discriminatory results, the need for transparency and explainability in AI decision-making to foster trust and support legal scrutiny, and intellectual property rights pertaining to AI algorithms and datasets (53). For AI to improve patient care in emergency medicine while reducing potential risks, legal and regulatory frameworks must change to keep up with the technology's advancements. These frameworks must address concerns of safety, ethics, and accountability (43)

Conclusions and Implications

AI, with its innovative capabilities to support triage, diagnosis, prognostication, and operational management, is rapidly transforming emergency medicine. AI tools are already being used in various ways, from clinical decision support systems and predictive analytics platforms to AI-supported emergency dispatch systems and patient-facing symptom checkers. According to preliminary assessments, these technologies have the potential to improve patient flow in emergency rooms, decrease the time it takes to treat critical conditions, and increase diagnostic accuracy, all of which can help address enduring issues like overcrowding, a lack of resources, and a heavy clinical burden. Although encouraging, the current body of evidence is still inconclusive. There have been little prospective validation or randomized clinical trials that verify improvements in patient outcomes, and the majority of published studies to date are retrospective. Premature or careless adoption of AI systems, if not thoroughly evaluated, risks unforeseen consequences, like missed diagnoses, skewed decision-making, or a decline in patient trust. If AI is to be used in emergency situations in a way that is both safe and moral, then key issues like algorithmic bias, lack of explanation, and variable integration with clinical workflows need to be carefully considered. Stakeholders in the clinical, research, ethical, and regulatory domains must consider several implications. First, to evaluate the practical effects of AI tools, there is an urgent need for thorough and methodologically sound research, especially multicenter prospective studies. Meaningful outcomes such as mortality, complication rates, patient satisfaction, and system-level efficiency should be included in these studies in addition to accuracy metrics. To identify deterioration over time and guarantee continued safety and relevance, systems for continuous performance monitoring and model recalibration should be put in place concurrently. Second, the deployment of AI must continue to revolve around ethical implementation. This means creating inclusive, equitable, and transparent systems with datasets that reflect a range of demographics and use cases. When implementing AI, hospitals should make sure that emergency personnel are properly trained in AI literacy and

that system design incorporates human oversight. Before new tools are put into use, institutional review boards or AI ethics committees may be crucial in ensuring that they are in line with ethical requirements and clinical needs. Third, it's crucial to navigate the constantly changing legal and policy landscape. Precise liability standards are required to maintain clinician trust and establish accountability in the event of AI-related harm. GDPR and HIPAA compliance, among other privacy laws, must be proactively incorporated into system architecture rather than being an afterthought. Crucially, legal frameworks ought to develop in tandem with technological advancements, providing safeguards without impeding responsible advancement. Last but not least, successful AI integration in emergency medicine will require interdisciplinary collaboration. To guarantee that AI systems are not only technically sound but also clinically beneficial, morally sound, and socially acceptable, clinicians, data scientists, engineers, ethicists, and legislators must collaborate from the very beginning of development. While working with legal experts can help anticipate and mitigate risks related to consent, bias, and accountability, involving frontline users in the design process can help ensure that tools are in line with the complex realities of the ED.

Conclusion

Considering all factors, AI holds immense potential to improve the promptness, precision, and fairness of emergency care. However, achieving this potential will require more than just technological advancement; it will also require our shared dedication to evidence-based development, ethical responsibility, and operational and legal preparedness. When used carefully, AI can be a strong ally of emergency physicians, complementing human knowledge rather than taking its place and ultimately helping to create a more secure, responsive, and compassionate emergency care system.

Ethics

Acknowledgement: The study was supported by the World Academic Council of Emergency Medicine (WACEM).

Footnotes

Author Contributions

Concept: T.B., I.S., L.S., Design: T.B., I.S., E.F.V., L.S., Data Collection or Processing: I.S., M.P., L.S., Analysis or Interpretation: T.B. I.S., E.F.V., M.P., F-X.D., A.L., L.S., Literature Search: T.B. I.S., M.P., F-X.D., A.L., L.S., Writing: T.B. I.S., E.F.V., M.P., F-X.D., A.L., L.S.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Chenais G, Lagarde E, Gil-Jardiné C. Artificial intelligence in emergency medicine: viewpoint of current applications and foreseeable opportunities and challenges. J Med Internet Res. 2023;25:e40031.
2. Kirubarajan A, Taher A, Khan S, Masood S. Artificial intelligence in emergency medicine: a scoping review. J Am Coll Emerg Physicians Open. 2020;1:1691-702.
3. Sackeim M. Banning abortion prevents us from providing safe care to all pregnant women. BMJ. 2024;387:q2459.
4. Wallace W, Chan C, Chidambaram S, Hanna L, Iqbal FM, Acharya A, et al. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. NPJ Digit Med. 2022;5:118.
5. Porto BM. Improving triage performance in emergency departments using machine learning and natural language processing: a systematic review. BMC Emerg Med. 2024;24:219.
6. Kopka M, Napierala H, Privoznik M, Sapunova D, Zhang S, Feufel MA. Evaluating self-triage accuracy of laypeople, symptom-assessment apps, and large language models: a framework for case vignette development using a representative design approach (RepVig). Cold Spring Harbor Laboratory. 2024.
7. Kachman MM, Brennan I, Oskvarek JJ, Waseem T, Pines JM. How artificial intelligence could transform emergency care. Am J Emerg Med. 2024;81:40-6.
8. Williams EL, Huynh D, Estai M, Sinha T, Summerscales M, Kanagasingam Y. Predicting inpatient admissions from emergency department triage using machine learning: a systematic review. Mayo Clin Proc Digit Health. 2025;3:100197.
9. Look CS, Teixayavong S, Djärv T, Ho AF, Tan KB, Ong ME. Improved interpretable machine learning emergency department triage tool addressing class imbalance. Digit Health. 20240;10:20552076241240910.
10. Blomberg SN, Folke F, Ersbøll AK, Christensen HC, Torp-Pedersen C, Sayre MR, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. Resuscitation. 2019;138:322-9.
11. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA Jr, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care. 2019;23:64.
12. Liu Z, Shu W, Liu H, Zhang X, Chong W. Development and validation of interpretable machine learning models for triage patients admitted to the intensive care unit. PLoS One. 2025;20:e0317819.
13. Klug M, Barash Y, Bechler S, Resheff YS, Tron T, Ironi A, et al. A gradient boosting machine learning model for predicting early mortality in the emergency department triage: devising a nine-point triage score. J Gen Intern Med. 2020;35:220-7.
14. Wang H, Sambamoorthi N, Hoot N, Bryant D, Sambamoorthi U. Evaluating fairness of machine learning prediction of prolonged wait times in emergency department with interpretable eXtreme gradient boosting. PLOS Digit Health. 2025;4:e0000751.
15. Joseph JW, Leventhal EL, Grossestreuer AV, Wong ML, Joseph LJ, Nathanson LA, et al. Deep-learning approaches to identify critically Ill patients at emergency department triage using limited information. J Am Coll Emerg Physicians Open. 2020;1:773-81.
16. Chonde SJ, Ashour OM, Nembhard DA, Kremer GEO. Model comparison in emergency severity index level prediction. Expert Systems with Applications. 2013;40:6901-9.

17. Lin P, Argon NT, Cheng Q, Evans CS, Linthicum B, Liu Y, et al. Disparities in emergency department prioritization and rooming of patients with similar triage acuity score. Acad Emerg Med. 2022;29:1320-8.
18. Fernandes M, Mendes R, Vieira SM, Leite F, Palos C, Johnson A, et al. Predicting intensive care unit admission among patients presenting to the emergency department using machine learning and natural language processing. PLoS One. 2020;15:e0229331.
19. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. PLoS One. 2018;13:e0201016.
20. Aydin ÖF. The potential role of artificial intelligence in emergency medicine and medical education. J Med Sci. 2024;5:180-1.
21. Chen MC, Huang TY, Chen TY, Boonyarat P, Chang YC. Clinical narrative-aware deep neural network for emergency department critical outcome prediction. J Biomed Inform. 2023;138:104284.
22. Yao LH, Leung KC, Tsai CL, Huang CH, Fu LC. A Novel deep learning-based system for triage in the emergency department using electronic medical records: retrospective cohort study. J Med Internet Res. 2021;23:e27008.
23. Roquette BP, Nagano H, Marujo EC, Maiorano AC. Prediction of admission in pediatric emergency department with deep neural networks and triage textual data. Neural Netw. 2020;126:170-7.
24. Lin JT, Chen CC. Point-of-care ultrasound for rapid diagnosis of massive pulmonary embolism in the emergency department: a case report. Ultrasound Med Biol. 2024;50(Suppl 1):30.
25. Hernandez Torres SI, Holland L, Winter T, Ortiz R, Amezcua KL, Ruiz A, Thorpe CR, et al. Real-time deployment of ultrasound image interpretation AI models for emergency medicine triage using a swine model. Technologies. 2025;13:29.
26. Cao A, Messak K, Hu G, Tsang M. Impact of emergency department triage on timely ecg acquisition. Can J Cardiol. 2021;37(Suppl):79-80.
27. Lazzari C, Montemerani S, Fabrizi C, Sacchi C, Belperio A, Fantacci M, et al. Pre-hospital point-of-care troponin: is it possible to anticipate the diagnosis? A preliminary report. Diagnostics (Basel). 2025;15:220.
28. Ivanov O, Molander K, Dunne R, Liu S, Brecher D, Masek K, et al. Detection of sepsis during emergency department triage using machine learning. Arxiv. 2022.
29. Busari A. Hybrid deep learning architectures for multimodal data fusion in healthcare diagnostics. Int J Sci Res Sci & Technology. 2024;11:271-80.
30. Xiao Y, Zhang J, Chi C, Ma Y, Song A. Criticality and clinical department prediction of ED patients using machine learning based on heterogeneous medical data. Comput Biol Med. 2023;165:107390.
31. Townsend BA, Plant KL, Hodge VJ, Ashaolu O, Calinescu R. Medical practitioner perspectives on AI in emergency triage. Front Digit Health. 2023;5:1297073.
32. Mahajan P, Mollen C, Alpern ER, Baird-Cox K, Boothman RC, Chamberlain JM, et al. An operational framework to study diagnostic errors in emergency departments: findings from a consensus panel. J Patient Saf. 2021;17:570-5.
33. Cardosi JD, Shen H, Groner JI, Armstrong M, Xiang H. Machine learning for outcome predictions of patients with trauma during emergency department care. BMJ Health Care Inform. 2021;28:100407.
34. Yan L, Zhang J, Chen L, Zhu Z, Sheng X, Zheng G, et al. Predictive value of machine learning for the risk of in-hospital death in patients with heart failure: a systematic review and meta-analysis. Clin Cardiol. 2025;48:e70071.
35. Wynants L, Smits LJM, Van Calster B. Demystifying AI in healthcare. BMJ. 2020;370:m3505.
36. Liu NT, Holcomb JB, Wade CE. Artificial intelligence–enabled prediction of emergency department crowding. JAMA Network Open. 2022;5:e2226450.
37. Sun Y, Heng BH, Tay SY, Seow E. Predicting hospital admissions at emergency department triage using routine administrative data. Acad Emerg Med. 2011;18:844-50.
38. Luo C, Islam MN, Sheils NE, Buresh J, Schuemie MJ, Doshi JA, et al. dPQL: a lossless distributed algorithm for generalized linear mixed model with application to privacy-preserving hospital profiling. J Am Med Inform Assoc. 2022;29:1366-71.
39. Selvan C, Anwar BH, Naveen S, Bhanu ST. Ambulance route optimization in a mobile ambulance dispatch system using deep neural network (DNN). Sci Rep. 2025;15:14232.
40. Kumar V, Goodarzian F, Ghasemi P, Chan FTS, Gupta N. Artificial intelligence applications in healthcare supply chain networks under disaster conditions. International Journal of Production Research. 2025;63:395-403.
41. Zhang Z, Wang Y, Sun L, Liu X. Improving emergency department efficiency with AI-driven resource management: a multicenter evaluation. JMIR Form Res. 2022;6:28199.
42. Smith ME, Zalesky CC, Lee S, Gottlieb M, Adhikari S, Goebel M, et al. Artificial intelligence in emergency medicine: a primer for the nonexpert. J Am Coll Emerg Physicians Open. 2025;6:100051.
43. Sharma N. Artificial intelligence: legal implications and challenges. Revista Opinião Jurídica. 2022;20:180-96.
44. Utomi E, Osifowokan AS, Donkor AA, Yowetu IA. Evaluating the impact of data protection compliance on AI development and deployment in the U.S. Health sector. World Journal of Advanced Research and Reviews. 2024;24:1100-10.
45. Gerke S, Rezaeikhonakdar D. Privacy aspects of direct-to-consumer artificial intelligence/machine learning health apps. Intelligence-Based Medicine. 2022;6:100061.
46. Mbah GO. Data privacy in the era of AI: navigating regulatory landscapes for global businesses. Int J Sci Res Arch. 2024;13:2040-58.
47. Casini P. Data protection in the European Union institutions from an information management perspective. in Recordkeeping in International Organizations; Routledge. 2020.p.28-58.
48. Ahluwalia M. Legal governance of brain data derived from artificial intelligence. Voices in Bioethics. 2021;7:1-5.
49. Moch E. Liability issues in the context of artificial intelligence: legal challenges andSolutions for AI-supported decisions. East African Journal of Law and Ethics. 2024;7:214-34.
50. Shumway DO, Hartman HJ. Medical malpractice liability in large language model artificial intelligence: legal review and policy recommendations. J Osteopath Med. 2024;124:287-90.
51. Rubisz S. Legal liability of an organisation using artificial intelligence. Scientific Papers of Silesian University of Technology Organization and Management Series. 2024;2024:493-507.
52. Okuno MJ, Okuno HG. Legal frameworks for AI service business participants: a comparative analysis of liability protection across jurisdictions. AI & SOCIETY. 2025.
53. Söderlund K, Larsson S. Enforcement design patterns in EU law: an analysis of the AI act. Digital Society. 2024;3:1-24.
54. Khan F. Regulating the revolution: a legal roadmap to optimizing AI in healthcare. SSRN Electronic Journal. 2023;25:50-71.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Case Report
DOI: 10.4274/eajem.galenos.2025.96992
Eurasian J Emerg Med. 2026;25: 29-33

Cerebral Fat Embolism Syndrome-case Report

Mehmet Berk Eyinç[1], Dilek Atik[2], Beril Cevizoğlu[1], Boran Polat[2], Habib Ali Yalama[2], Ömer Faruk Demirayak[1]

[1]Karaman Training and Research Hospital, Clinic of Emergency Department, Karaman, Türkiye
[2]Karamanoğlu Mehmetbey University Faculty of Medicine, Department of Emergency Medicine, Karaman, Türkiye

Abstract

Fat embolism syndrome (FES) is an uncommon complication, typically presenting with respiratory insufficiency following orthopedic trauma. The condition is referred to as cerebral FES (CFES) when neurological symptoms dominate, which is an even rarer form of FES. We present a case of a 21-year-old male who was brought to the emergency department with lower extremity fractures following a traffic accident. The patient experienced a sudden decline in consciousness and developed confusion during his follow-up. An echocardiogram performed in the emergency room revealed hyperechogenic structures in heart chambers, and subsequent magnetic resonance imaging scan showed multiple infarct areas, leading to a diagnosis of CFES. The diagnosis of CFES is challenging due to the non-specific nature of the symptoms and the frequent normality of computed tomography scans. High clinical suspicion is crucial in patients with orthopedic injuries who experience sudden neurological decline.

Keywords: Fat embolism, orthopedic trauma, cerebral fat embolism syndrome

Introduction

Fat embolism syndrome (FES) is a rare complication, typically presenting with respiratory insufficiency following orthopedic trauma (1). Diagnostic criteria for FES were first established by Gurd in 1970 and later modified by Wilson, becoming widely used. He proposed that the diagnosis of FES be established by the simultaneous identification of at least two major clinical findings or one major and four minor clinical findings. The major criteria encompass respiratory insufficiency, neurological disturbances, and petechial rash. Minor criteria include the detection of fat globules in urine or sputum, renal dysfunction, jaundice, retinal changes, pyrexia, tachycardia, unexplained alterations in platelet (PLT) count or hematocrit, and the presence of fat macroglobulinemia (2,3). Other classifications include Lindeque's classification, which focuses on respiratory symptoms, and Schonfeld's classification, which develops a scoring system. The Schonfeld classification is organized so that Petechiae are 5 points, Chest X-ray changes (diffuse alveolar infiltrates) are 4 points, hypoxemia is 3 points, fever (>38 °C), tachycardia (>120 bpm), tachypnea (>30 bpm), and confusion are each given one point (2,4). Regardless of the classification system, the classic presentation includes a triad of respiratory distress, petechial rash, and neurological changes (3). When neurological symptoms dominate, the condition is referred to as cerebral FES (CFES) (5). Early recognition is crucial, especially in orthopedic trauma patients showing sudden neurological decline. Diagnosing CFES is challenging due to the lack of universal criteria, relying instead on clinical assessment and the patient's medical history (6).

Case Report

A 21-year-old male was brought to the emergency medicine department complaining of back and leg pain after a traffic accident. There is no history of drug use other than psychiatric medication. Upon arrival at the emergency depeartment, his vital signs were stabile and glaskow coma scale is scrored 15. conscious, oriented-cooperative, pupillar isochoric, infrared +/+. Abdominal and respiratory examinations are normal. No obvious pathology was observed in the cranial nerve examination. Multiple bone fractures were present in the right lower extremity; muscle strength was not clearly assessed; no obvious muscle strength loss was observed in other extremities. The patient's imaging shows right femur, fibula and tibial shaft fractures (Figure 1), lung

Corresponding Author: Boran Polat MD, Karamanoğlu Mehmetbey University Faculty of Medicine, Department of Emergency Medicine, Karaman, Türkiye
E-mail: borannpolat@gmail.com **ORCID ID:** orcid.org/0009-0006-5436-8849
Cite this article as: Eyinç MB, Atik D, Cevizoğlu B, Polat B, Yalama HA, Demirayak ÖF. Cerebral fat embolism syndrome-case report. Eurasian J Emerg Med. 2026;25: 29-33.

Received: 25.06.2025
Accepted: 03.09.2025
Epub: 23.10.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

contusion, and C6 corpus fracture. No pathological findings were observed in abdominal imaging. No pathological findings were present in the laboratory results obtained during admission to the hospital. Toxicology results were normal. Laboratory data showed a drop in hemoglobin (HGB) during its follow-up at emergency department (HGB: 15.1 g/dL, first HGB: 16.6 g/dL). Also, during his intensive care unit stay, he was reported to have a petechial rash, mostly on the anterior surface of his trunk. There was a decrease in PLT count (PLT: 85K/uL) and a drop in HGB (HGB: 10.1 g/dL, after orthopedic surgery HGB: 14.3 g/dL, reference range: 12-18 g/dL).

The patient, who did not need urgent surgery by the thoracic and neurosurgery departments, had a long-leg splint applied to his right lower extremity and was monitored according to the operation plan, with a Philadelphia neck collar attached. During the follow-up, the patient's vital signs remained stable. Approximately 11-12 hours after admission, he suddenly developed confusion and agitation followed by a rapid decline in the level of consciousness. Thereupon, an emergency bedside ultrasound was performed. Upon the detection of thrombus-like structures at Echo (Figure 2), computed tomography (CT) and diffusion magnetic resonance imaging (MRI) were performed. According to diffusion MRI (Figure 3), focal areas were observed in the localization of both basal ganglia, some of which restrict linear moderate diffusion, and these areas were in the millimeter dimension. In addition, a focal area restricting diffusion in millimeter dimensions was observed in the right half of the pons. Focal area restricting linear moderate diffusion at the

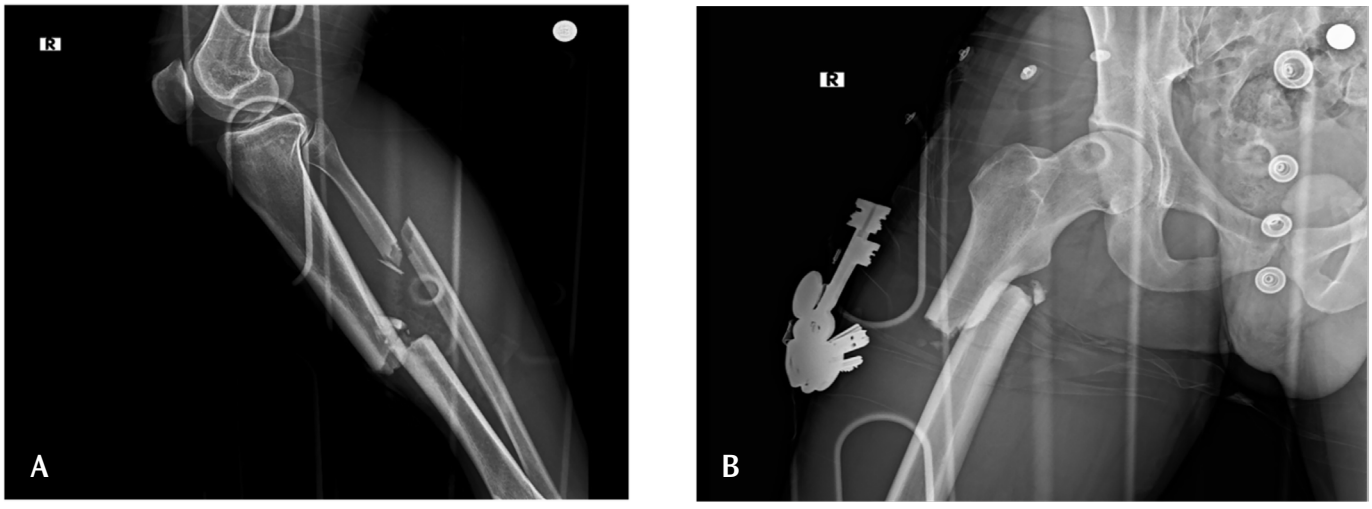


Figure 1. (A) Femur shaft fracture, (B) Fibula and tibial shaft fractures

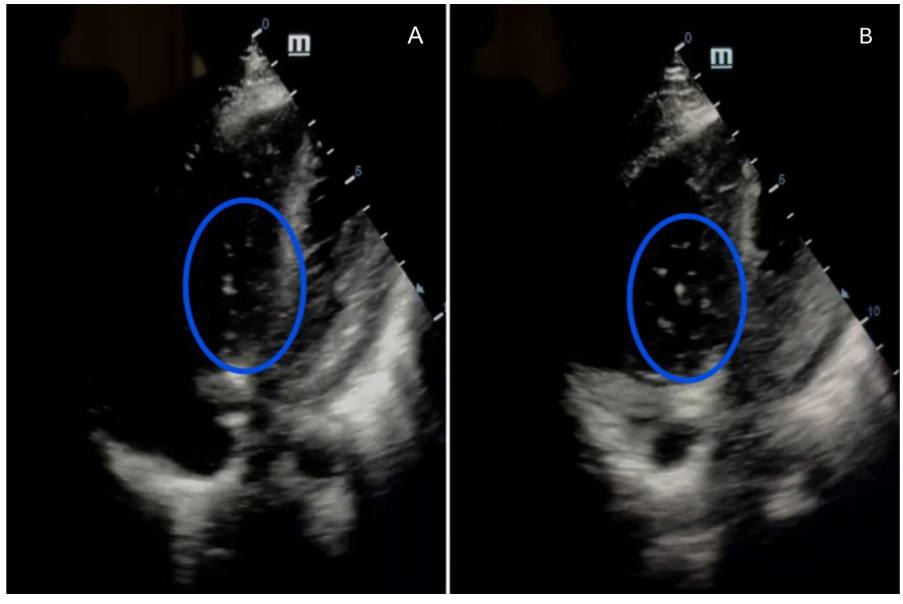


Figure 2. (A, B) Thrombus-like structures highlighted by blue circles on echocardiography

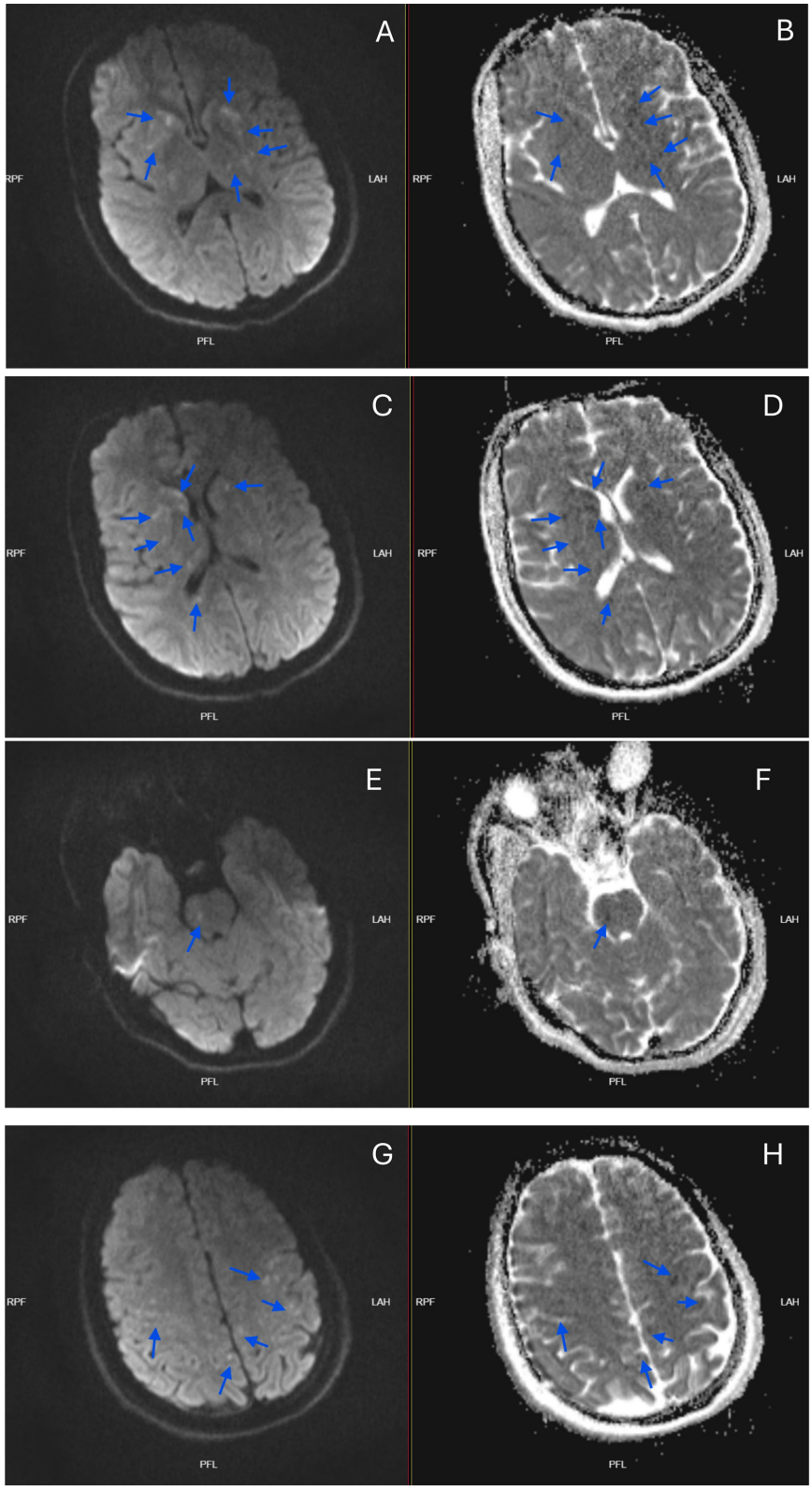


Figure 3. Multiple infarct areas on MRI (starfield pattern). Axial diffusion-weighted imaging (DWI) and corresponding apparent diffusion coefficient images (A–H), demonstrating areas of diffusion restriction indicated by blue arrows

MRI: Magnetic resonance imaging, RPF: Right posterior frontal, PFL: Posterior fronto-lateral, LAH: Left anterior hemisphere

cortical level was observed at the centrum semiovale level, and focal areas restricting diffusion in millimeter size in the frontal and parietal lobes on the left were also observed. MRI findings are consistent with a starfield pattern suggestive of cerebral fat embolism. There is no hemorrhage on CT. Additionally, the patient developed desaturation and underwent a pulmonary CT angiography to rule out pulmonary pathologies, but pulmonary embolism was detected. According to the report, a smear-like effusion was observed in both hemithoraxes, and nonspecific density increases were observed in the dependent segments. With supportive treatment; intensive care follow-up was planned with a preliminary diagnosis of fat embolism.

Discussion

FES is an uncommon disorder primarily characterized by respiratory insufficiency, often arising as a complication following orthopedic trauma (1). Typically presenting within 1-2 days following trauma involving long-bone fractures or orthopedic surgery, it manifests as a petechial rash, progressive respiratory insufficiency, and declining mental status (7). Only a small subset of patients, comprising 0.9% to 2.2%, subsequently develops FES, which is classically characterized by a triad of respiratory symptoms, petechial rash, and alterations in mental status (6). The diagnosis of FES is predominantly clinical; however, the Gurd and Wilson criteria, as well as the Schonfeld fat embolism index, remain two of the most widely utilized and validated diagnostic frameworks (4). In our case, the patient fulfilled 3 major criteria of the Gurd and Wilson Criteria for FES: respiratory distress and petechial rash and neurological symptoms as altered mental status, along with minor criterias, drop in HGB (from 16.6 to 10.1 g/dL), thrombocytopenia (85 K/uL) and according to the Schonfeld fat embolism index, the patient met the following criteria: petechial rash (5 points), and confusion(1 point).

The term "CFES" is used to describe cases where neurological symptoms are the predominant features (5). CFES is a rare variant of FES characterized by isolated neurologic symptoms including ischemic and hemorrhagic strokes, seizures, convulsive and non-convulsive status epilepticus, autonomic impairment, acute encephalopathy, and coma (1,6). Literature reviews and case reports have shown that CFES is common in men with bilateral femur fractures in their early 30s (8).

This diagnosis should be considered in trauma patients with long bone fractures who were found to have confusion or focal neurologic deficits. CFE should also be considered in patients with a history of remote trauma, as in this case, who present with new-onset neurologic symptoms (8).

The hypothesized mechanism is that fat globules enter arterial circulation either through an intracardiac shunt such as a patent foramen ovale, or, more commonly, micro globules filter directly through lung capillaries to reach arterial circulation (6).

MRI utilizing either diffusion-weighted imaging or susceptibility-weighted imaging sequences offers the greatest sensitivity for confirming the diagnosis (9,10). While there are no pathognomonic imaging findings, the "starfield pattern" observed on MRI is the most prevalent and recognized manifestation of CFES (11). This imaging finding is also consistent with the MRI findings in our case, further supporting the possibility of CFES in our patient.

Also, in the study by Maghrebi et al. (12), transthoracic echocardiography showed the "snowstorm" appearance of numerous hyperechoic particles in the inferior vena cava. Bedside echocardiographic evaluation in our patient demonstrated hyperechoic structures suggestive of thrombus formation. There are no definitive or specific treatments of FES; therefore, management is entirely supportive (9).

Considering the multifactorial etiology of FES, early operative fixation of long bone fractures is argued to reduce the incidence of FES, although it cannot be prevented in all patients (13,14).

In our case, we discussed a patient who was admitted to the hospital as a result of a traffic accident and who developed a change in consciousness during follow-up. Based on the clinical, laboratory, and imaging findings, differential diagnoses, such as diffuse axonal injury, disseminated intravascular coagulation, and intraparenchymal, subdural, or epidural hemorrhages, should be considered in trauma patients presenting with altered mental status. However, cerebral fat embolism was considered the more likely diagnosis in our case.

Conclusion

The diagnosis of CFES is challenging due to the non-specific nature of the symptoms and the frequent normal results of CT scans. High suspicion should be maintained in patients with orthopedic injuries who exhibit sudden neurological decline. There are no universal diagnostic criteria for CFES, making the diagnosis dependent on the patient's medical history and clinical manifestations (6).

Ethic

Informed Consent: Written informed consent was obtained from the patient for publication of this case and related images.

Footnotes

Author Contributions

Surgical and Medical Practices: D.A., B.P., H.A.Y., Ö.F.D., Concept: D.A., B.P., Design: D.A., H.A.Y., Data Collection or Processing: M.B.E., B.C., Analysis or Interpretation: M.B.E., B.C., Literature Search: M.B.E., Writing: M.B.E., B.P.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Morales-Vidal SG. Neurologic complications of fat embolism syndrome. Curr Neurol Neurosci Rep. 2019;19:14.
2. Kwon J, Coimbra R. Fat embolism syndrome after trauma: what you need to know. J Trauma Acute Care Surg. 2024;97:505-13.
3. Morena D, Scopetti M, Padovano M, Turillazzi E, Fineschi V. Fat embolism: a systematic review to facilitate the development of standardised procedures in pathology. Histopathology. 2025;86:845-61.
4. Gholipour M, Salimi M, Motamedi A, Abbasi F. Fat embolism syndrome following femoral shaft fracture: a case report and diagnostic considerations. Radiol Case Rep. 2025;20:943-8.
5. Singh A, Davis AP, Taylor B, Peters S, Tirschwell D, Longstreth WT, et al. Cerebral fat embolism syndrome at a single trauma center. J Stroke Cerebrovasc Dis. 2022;31:106794.
6. Saran JS, Hussain AH, Papadakos PJ, Roberts D. Cerebral fat embolism syndrome. J Emerg Med. 2020;58:95-6.
7. Godoy DA, Di Napoli M, Rabinstein AA. Cerebral fat embolism: recognition, complications, and prognosis. Neurocrit Care. 2018;29:358-65.
8. Gibbons P, Kennedy M, Borczuk P, Baymon D. Cerebral fat embolism after a fall. J Emerg Med. 2022;63:87-90.
9. Fukumoto LE, Fukumoto KD. Fat embolism syndrome. Nurs Clin North Am. 2018;53:335-47.
10. Kuo KH, Pan YJ, Lai YJ, Cheung WK, Chang FC, Jarosz J. Dynamic MR imaging patterns of cerebral fat embolism: a systematic review with ıllustrative cases. AJNR Am J Neuroradiol. 2014;35:1052-7.
11. Parizel PM, Demey HE, Veeckmans G, Verstreken F, Cras P, Jorens PG, et al. Early diagnosis of cerebral fat embolism syndrome by diffusion-weighted MRI (starfield pattern). Stroke. 2001;32:2942-4.
12. Maghrebi S, Cheikhrouhou H, Triki Z, Karoui A. Transthoracic echocardiography in fat embolism: a real-time diagnostic tool. J Cardiothorac Vasc Anesth. 2017;31:47-8.
13. Blokhuis TJ, Pape HC, Frölke JP. Timing of definitive fixation of major long bone fractures: can fat embolism syndrome be prevented? Injury. 2017;48:3-6.
14. Kwiatt ME, Seamon MJ. Fat embolism syndrome. Int J Crit Illn Inj Sci. 2013;3:64-8.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.65390
Eurasian J Emerg Med. 2026;25: 34-7

Retrospective Analysis: The Effect of the Seasonal Changes on the Frequency of Urinary System Stone Operations

Bayram Doğan[1], Selin Ayten Akok[1], Abdullah İlktaç[1], Mehmet Emre Demirbilek[2], Fatih Gevher[1], Serkan Akıncı[1], Yusuf Özlem İlbey[1]

[1]Bezmialem Vakıf University Faculty of Medicine, Department of Urology, İstanbul, Türkiye
[2]University of Health Sciences Türkiye, Şişli Hamidiye Etfal Training and Research Hospital, Clinic of Dermatology and Venereology, İstanbul, Türkiye

Abstract

Aim: Urolithiasis is a multifactorial disease influenced by various factors, including climate and seasonal changes. While climatic effects on stone formation have been studied, seasonal variations in urinary stone surgeries remain underexplored. To investigate the seasonal variation in the frequency of urinary stone surgeries over a three-year period and assess potential seasonal influences on surgical incidence and emergency department visits.

Materials and Methods: A retrospective review of 841 urinary stone surgeries and 509 transurethral resections of the prostate (TUR-P) procedures (as controls) performed between January 2018 and December 2020 was conducted. Procedures were categorized by meteorological seasons. Statistical analysis was performed using chi-square tests to evaluate differences in seasonal distribution.

Results: Urinary stone surgeries showed a significant seasonal variation, with the highest frequency in summer (27.8%) and the lowest in spring (19.3%) ($p<0.05$). No similar seasonal trend was observed in TUR-P procedures. The majority of patients were female (66.0%) with a mean age of 49.7 years. The ureter and kidneys were the most common locations of stone formation. Most patients underwent ureterorenoscopy, and postoperative DJ stenting was frequent (79.5%). Residual stones and complications were infrequent.

Conclusion: The findings demonstrate a statistically significant increase in urinary stone surgeries during summer, suggesting seasonal influences on disease manifestation. These results highlight the importance of heightened clinical awareness and preventive measures, particularly during warmer months, to reduce the burden of urinary stone disease and emergency department visits.

Keywords: Urolithiasis, seasonal variation, urinary stone surgery, summer incidence, hydration, endourological procedures

Introduction

Urolithiasis is a multifactorial disease characterized by the formation of calculi in the kidney, ureter, bladder, or urethra. It is a common cause of emergency department visits, most frequently presenting as renal colic; an acute and severe flank pain that often radiates to the groin and may be accompanied by hematuria and dysuria. The prevalence of renal colic varies globally, ranging between 5% and 15%, depending on geographical, environmental, genetic, and lifestyle-related factors (1). Recurrence rates are as high as 50%, and the incidence is significantly higher in men (2).

Numerous factors contribute to the development of urinary stones, including race, gender, body mass index, dietary patterns, fluid intake, and climate. Certain "stone belt" regions have been identified at regional and global levels, where the incidence of urolithiasis is disproportionately high. For instance, populations in the Western Hemisphere demonstrate elevated prevalence rates (approximately 5-9% in Europe, 12% in Canada, and 13-15% in the United States, particularly in the southeastern states) compared to those in the Eastern Hemisphere, where rates are generally lower (1-5%) (3,4). A well-known high-risk region is the so-called Afro-Asian stone belt (5), which spans from Egypt, Saudi

Corresponding Author: Selin Ayten Akok MD, Bezmialem Vakıf University Faculty of Medicine, Department of Urology, İstanbul, Türkiye
E-mail: selinayten@gmail.com **ORCID ID:** orcid.org/0009-0003-4847-5945
Cite this article as: Doğan B, Akok SA, İlktaç A, Demirbilek ME, Gevher F, Akıncı S, et al. Retrospective analysis: the effect of the seasonal changes on the frequency of urinary system stone operations. Eurasian J Emerg Med. 2026;25: 34-7.

Received: 16.08.2025
Accepted: 03.09.2025
Epub: 23.10.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Arabia, Sudan and the UAE to Kuwait, Iran, Pakistan, Thailand, India and the Philippines, with prevalence figures ranging from 4% to 20% in various studies (6). Climatic and seasonal factors have gained increasing attention due to their potential impact on hydration status and metabolic processes involved in stone formation. In hot and arid environments, increased perspiration and insensible water loss lead to reduced urinary volume and increased solute concentration, which promote crystal formation. Moreover, greater exposure to sunlight stimulates vitamin D synthesis, which may result in hypercalcemia—a known risk factor for calcium-based stones. These mechanisms suggest that urinary stone incidence may vary with seasonal changes, particularly in warmer months when dehydration is more likely.

Although the relationship between climate and urolithiasis has been previously examined, few studies have focused specifically on seasonal variation in the frequency of stone-related surgical interventions. Moreover, much of the existing literature on this topic is several decades old and may not accurately reflect current population dynamics or environmental trends. The primary aim of this study is to determine whether a seasonal pattern exists in the surgical management of urolithiasis by comparing the frequency of urinary stone operations across different seasons. The findings may help inform preventive strategies, guide future etiologic research, and provide insights into potential seasonal influences on clinical practice.

Materials and Methods

Between 1 January 2018 and 31 December 2020, a total of 841 urinary stone surgeries and 509 transurethral resections of the prostate (TUR-P) procedures—selected as the control group—were retrospectively reviewed and classified according to the four meteorological seasons: spring (March–May), summer (June–August), autumn (September–November), and winter (December–February). TUR-P cases were chosen as the control group based on the assumption that, unlike urinary stone disease, benign prostatic hyperplasia and its surgical treatment are not influenced by environmental or seasonal factors. This comparison aimed to determine whether the incidence of urinary stone surgeries demonstrated a seasonal trend beyond general surgical activity. This study was reviewed and approved by the Ethics Committee of Bezmialem University for the application titled "Retrospective Analysis: The Effect of Seasonal Changes on the Frequency of Urinary System Stone Surgeries" during the 11th Interventional Research Committee meeting, with unanimous approval (decision number: E-54022451-050.05.04-19936, date: 25.05.2021). All procedures were conducted in accordance with the principles of the Declaration of Helsinki.

Statistical Analysis

Data were analyzed using IBM SPSS Statistics, version 22.0, a widely used software package in biomedical research. The primary statistical method for categorical variables was the chi-square test, used to assess associations between groups and determine whether observed differences were statistically significant. A p value of <0.05 was considered indicative of statistical significance.

Results

A total of 841 urinary stone operations performed between January 1, 2018, and December 31, 2020, were evaluated. The seasonal distribution of these procedures was as follows: 19.3% (n=162) in spring, 27.8% (n=234) in summer, 26.3% (n=221) in autumn, and 26.6% (n=224) in winter. In the control group, consisting of 509 patients who underwent TUR-P during the same period, the seasonal distribution was 15.9% (n=81) in spring, 22.2% (n=113) in summer, 29.7% (n=151) in autumn, and 32.2% (n=164) in winter. According to the comparative analysis, the rate of urinary stone operations was significantly higher in summer than in spring (27.8% vs. 19.3%, $p<0.05$). Additionally, the rate of urinary stone operations in summer was significantly higher than the rate of TUR-P operations in the same season (27.8% vs. 22.2%, $p<0.05$). The association between surgical procedure type and seasonal distribution was analyzed using the Pearson chi-square test, indicating a statistically significant difference in seasonal distributions between the two groups ($\chi^2=10.640$, $p=0.014$).

The mean age of the study population was 49.7±15.8 years, with a predominance of female patients (66.0%). The most common stone locations were the ureter (45.9%) and the kidney (44.1%). The majority of patients underwent Type 3 surgery (83.8%), followed by Type 4 (8.7%). Left-sided stones (54.0%) were slightly more frequent than right-sided stones (46.0%) among patients with laterality data (n=767). Hydronephrosis was present in 54.8% of the patients at diagnosis. Flank pain (75.1%) was the most common presenting symptom, followed by nausea/vomiting (5.5%) and urinary retention (5.5%). More than half of the patients (58.1%) had a history of previous stone disease, and 27.5% had undergone prior stone surgery. Preoperative extracorporeal shock wave lithotripsy (ESWL) was reported in 7.6%, while only 0.4% required postoperative ESWL. Residual stones were observed in 8.3% of the cases. A preoperative DJ stent was placed in 12.6% of patients, whereas postoperative DJ stents were used in 79.5%. Comorbidities were present in 34.6% of patients. Postoperative complications (fever, bleeding, infection, organ injury) were rare, occurring in only 1.1% of the patients. The mean stone size was

212.5±257.8 mm², and the mean creatinine level was 1.12±0.81 mg/dL. The mean symptom duration prior to presentation was 16.4±12.9 days, and the mean hospital stay was 2.05±2.54 days (Table 1).

Discussion

In this study, patients who underwent surgery for urinary system stones in our clinic over approximately three years, were retrospectively analysed to investigate whether seasonal temperature variations influence the frequency of urinary stone surgeries. Our results demonstrated a statistically significant increase in urinary stone surgeries during the summer months, with a corresponding decline in spring. In contrast, the seasonal distribution of TUR-P procedures, selected as a control group, remained relatively stable and was not significantly elevated in summer. These findings suggest a potential association between seasonality and the incidence of symptomatic or obstructive urinary stone disease requiring surgical intervention.

The observed summer peak in stone surgeries is consistent with existing evidence indicating that warm weather and dehydration may play a crucial role in stone formation and symptomatic presentation (7). Several pathophysiological mechanisms may explain this relationship. High ambient temperatures during summer months increase insensible water loss through perspiration, leading to reduced urine output and increased urinary solute concentration—key factors in promoting crystallization and stone formation. Additionally, enhanced sunlight exposure may lead to increased vitamin D synthesis and calcium absorption, potentially raising urinary calcium levels and promoting calcium-based stones (8,9). Geographical studies have similarly identified higher stone disease prevalence in hot, arid regions, often referred to as “stone belts”. Our findings support this climate-stone link and extend it to surgical outcomes, emphasizing that environmental factors may not only influence stone formation but also the clinical burden on healthcare systems during warmer months. It is noteworthy that despite global warming and climate shifts, seasonal patterns in stone-related hospital visits and procedures remain evident, reinforcing the relevance of weather-related risk factors.

The demographic and clinical characteristics of our study population provide additional context for the observed seasonal variation. The majority of the patients were female, with a mean age of approximately 50 years. While stone disease has traditionally been more common in men, our data showed a higher proportion of female patients. This finding may reflect changing dietary and lifestyle patterns, as recent studies suggest a narrowing of the historical sex disparity in stone disease prevalence (10). Flank pain was the most frequently reported

Table 1. Demographic and clinical characteristics of patients undergoing urinary stone surgery

Variable	Value
Age (years), mean ± SD	49.7±15.8
Sex, n (%)	
Male	286 (34.0%)
Female	555 (66.0%)
Stone location, n (%)	
Renal	371 (44.1%)
Ureteral	386 (45.9%)
Bladder	73 (8.7%)
Urethral	11 (1.3%)
Type of operation, n (%)	
Percutaneous nephrolithotomy	59 (7.0%)
Open nephrolithotomy	4 (0.5%)
Ureterorenoscopy	705 (83.8%)
Cystolithotripsy	73 (8.7%)
Laterality (n=767), n (%)	
Right	353 (46.0%)
Left	414 (54.0%)
Hydronephrosis, n (%)	
Present	461 (54.8%)
Absent	380 (45.2%)
Presenting symptom, n (%)	
Flank pain	632 (75.1%)
Hematuria	40 (4.8%)
Nausea/vomiting	46 (5.5%)
Dysuria	36 (4.3%)
Urinary retention	46 (5.5%)
Others	41 (4.8%)
Previous stone history, n (%)	
Yes	489 (58.1%)
No	352 (41.9%)
Previous stone surgery, n (%)	
Yes	231 (27.5%)
No	610 (72.5%)
Residual stone, n (%)	
Yes	70 (8.3%)
No	771 (91.7%)
Preoperative DJ stent, n (%)	
Yes	106 (12.6%)
No	735 (87.4%)
Postoperative DJ stent, n (%)	
Yes	669 (79.5%)
No	172 (20.5%)
Comorbidities, n (%)	
None	550 (65.4%)
≥1 comorbidity	291 (34.6%)
Postoperative complication, n (%)	
Yes	9 (1.1%)
No	832 (98.9%)
Stone size (mm²), mean ± SD	212.5±257.8
Symptom duration (days), mean ± SD	16.4±12.9
Hospital stay (days), mean ± SD	2.05±2.54

SD: Standard deviation

presenting symptom, and the most common stone locations were the ureter and kidney. More than half of the patients (58.1%) had a history of previous stone disease, consistent with the high recurrence rates reported in the literature. Therefore, it is recommended that individuals with urinary stone disease undergo regular follow-up visits in the urology outpatient clinic at specified intervals. Most patients underwent ureterorenoscopy, and over 79% received a postoperative DJ stent, reflecting a preference for endourological management strategies in current clinical practice. Residual stones and postoperative complications were infrequent, suggesting effective perioperative planning and surgical proficiency (11).

Study Limitations

Despite the robustness of the dataset, this study has several limitations. First, its retrospective design precludes establishing causality. Second, meteorological variables such as temperature, humidity, and sunlight exposure were not included in the analysis, which could have provided more detailed insights into seasonal effects. Additionally, the study was conducted at a single institution, potentially limiting the generalizability of the findings to regions with different climates, healthcare systems, or patient demographics. Future research incorporating environmental data, multi-center collaboration, and prospective designs may yield a more comprehensive understanding of the seasonal dynamics of urinary stone disease. Moreover, preventive strategies—such as hydration education during high-risk seasons—should be evaluated for their potential to reduce symptomatic presentations and surgical burden.

Conclusion

In conclusion, our study found a statistically significant increase in urinary stone surgeries during summer months, suggesting a seasonal influence on disease manifestation. These findings underscore the need for heightened clinical awareness and preventive strategies during warmer seasons to reduce the morbidity and healthcare resource utilization associated with urinary stone disease.

Ethics

Ethics Committee Approval: The study was approved by the Ethics Committee of Bezmialem University (decision number: E-54022451-050.05.04-19936, date: 25.05.2021). All procedures were conducted in accordance with the principles of the Declaration of Helsinki.

Informed Consent: This is a retrospective study.

Footnotes

Author Contributions

Surgical and Medical Practices: B.D., A.İ., F.G., S.A., Y.Ö.İ., Concept: B.D., Design: B.D., A.İ., F.G., Data Collection or Processing: M.E.D., Analysis or Interpretation: M.E.D., Literature Search: S.A.A., F.G., S.A., Writing: B.D., S.A.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Fisang C, Anding R, Müller SC, Latz S, Laube N. Urolithiasis--an interdisciplinary diagnostic, therapeutic and secondary preventive challenge. Deutsches Arztebl Int. 2015;112:83-91.
2. Xu JZ, Li C, Xia QD, Lu JL, Wan ZC, Hu L, et al. Sex disparities and the risk of urolithiasis: a large cross-sectional study. Ann Med. 2022;54:1627-35.
3. Ghani KR, Roghmann F, Sammon JD, Trudeau V, Sukumar S, Rahbar H, et al. Emergency department visits in the United States for upper urinary tract stones: trends in hospitalization and charges. J Urol. 2014;191:90-6. Epub 2013 Aug 7.
4. Elder JW, Delgado MK, Chung BI, Pirrotta EA, Wang NE. Variation in the intensity of care for patients with uncomplicated renal colic presenting to U.S. emergency departments. J Emerg Med. 2016;51:628-35. Epub 2016 Oct 5.
5. López M, Hoppe B. History, epidemiology and regional diversities of urolithiasis. Pediatr Nephrol. 2010;25:49-59.
6. Rizvi SA, Naqvi SA, Hussain Z, Hashmi A, Hussain M, Zafar MN, et al. The management of stone disease. BJU Int. 2002;89(Suppl 1):62-8.
7. Tasian GE, Pulido JE, Gasparrini A, Saigal CS, Horton BP, Landis JR, et al. Urologic Diseases in America Project. Daily mean temperature and clinical kidney stone presentation in five U.S. metropolitan areas: a time-series analysis. Environ Health Perspect. 2014;122:1081-7. Epub 2014 Jul 10.
8. Sohgaura A, Bigoniya P. A review on epidemiology and etiology of renal stone. Int J Pharm Pharm Sci. 2017;7:54-62.
9. Borghi L, Meschi T, Amato F, Novarini A, Giannini A, Quarantelli C, et al. Urinary volume, water and recurrences in idiopathic calcium nephrolithiasis: a 5-year randomized prospective study. J Urol. 1996;156:1613-9.
10. Gillams K, Juliebø-Jones P, Juliebø SØ, Somani BK. Gender differences in kidney stone disease (KSD): findings from a systematic review. Curr Urol Rep. 2021;22:50.
11. De Coninck V, Keller EX, Somani B, Giusti G, Proietti S, Rodriguez-Socarras M, et al. Complications of ureteroscopy: a complete overview World J Urol. 2020;38:2147-66. Epub 2019 Nov 20.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.25902
Eurasian J Emerg Med. 2026;25: 38-44

Evaluation of Survival Rates and Associated Factors After Cardiopulmonary Resuscitation

Mehmet Koşargelir, Güleser Akpınar, Abdullah Yaser Güney, Kenan Ahmet Türkdoğan

University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye

Abstract

Aim: The aim of this study was to determine survival rates in patients who underwent cardiopulmonary resuscitation (CPR) and to evaluate related demographic characteristics, clinical findings, and laboratory results.

Materials and Methods: Data from 620 patients who were transported by ambulance to the emergency department while receiving CPR were retrospectively analysed. Demographic characteristics, clinical findings, laboratory values, and time variables were recorded. Factors associated with survival were evaluated using the chi-square test, independent samples t-test, correlation analysis, and logistic regression analysis.

Results: The median age of the 620 patients included in the study was 67 years (IQR: 54-79), and 64.52% (n=400) were male. 95.32% (n=591) of the calls received by the emergency call centre originated from the region. 98.55% of patients (n=611) lived in urban areas. The one-month survival rate after CPR was found to be 12.42%. Successful resuscitation was achieved in 7.90% of these patients (n=49). In multivariate logistic regression analysis, pH ≤6.817 (OR: 37.39, 95% CI: 14.48-96.52, p<0.001), PO_2 (OR: 0.990, 95% CI: 0.984-0.996, p=0.002), platelet count (OR: 0.99, 95% CI: 0.99-1.00, p=0.034), neutrophil-to-lymphocyte ratio (OR: 0.83, 95% CI: 0.72-0.94, p=0.004), basophil count (OR: 0.18, 95% CI: 0.06-0.55, p=0.002), and MCHC (OR: 1.01, 95% CI: 1.01-1.01, p<0.001) were identified as significant independent predictors of mortality. The combined model incorporating pH and PO_2 demonstrated excellent discriminative ability with an AUC of 0.923 (95% CI: 0.875-0.971), sensitivity of 97.8%, and specificity of 83.1% at a cut-off probability of 0.700.

Conclusion: This study demonstrated that arterial blood gas findings (particularly pH ≥6.817 and PO_2 ≤45.5 mmHg) and certain hematological markers (platelet, neutrophil-to-lymphocyte ratio, basophil) have high diagnostic value in predicting mortality after cardiac arrest.

Keywords: Cardiopulmonary resuscitation, survival, platelet/lymphocyte ratio, neutrophil/lymphocyte ratio, predictive factors

Introduction

Cardiac arrest is defined as the cessation of systemic circulation following the termination of the heart's mechanical activity (1). Cardiopulmonary resuscitation (CPR) is a vital intervention in cardiac arrest and increases the likelihood of survival when performed successfully (2). However, post-CPR survival rates vary worldwide and are often below the desired level (3).

The literature reports survival rates of 8% after out-of-hospital cardiac arrest in European countries, while this rate reaches 10-12% in the United States. A study conducted in the United States in the case of in-hospital cardiac arrests found that the survival rate was approximately 25% (4,5). Recent literature reports that these rates have improved over the years (6,7). A study conducted in Türkiye found that the survival rate after out-of-hospital cardiac arrest was 6.9% (8).

Identifying the factors that influence survival rates is crucial for improving the effectiveness of CPR, various factors that may affect survival after cardiac arrest have been identified. These include age, the location and time of cardiac arrest, the occurrence of ventricular fibrillation, and ventricular tachycardia defined as shockable rhythms, diagnosis of cardiac arrest, the time to initiation of basic life support, and the time to defibrillation (9,10).

Corresponding Author: Güleser Akpınar MD, University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye
E-mail: guleserakpinarduman@gmail.com **ORCID ID:** orcid.org/0000-0001-8559-5098

Cite this article as: Koşargelir M, Akpınar G, Güney AY, Türkdoğan KA. Evaluation of survival rates and associated factors after cardiopulmonary resuscitation. Eurasian J Emerg Med. 2026;25: 38-44.

Received: 07.08.2025
Accepted: 09.09.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Recently, the effect of hematological parameters and inflammatory markers on prognosis after cardiac arrest has also been investigated (11-13). Hematological parameters such as the neutrophil-to-lymphocyte ratio (NLR) and platelet-lymphocyte ratio (PLR), which are indicators of the systemic inflammatory response, have been reported to have prognostic value in various cardiovascular diseases (14,15). These parameters are gaining importance in clinical practice as low-cost, easily accessible biomarkers that can be calculated from routine laboratory tests.

The pathophysiological process following cardiac arrest is complex and involves many components, including ischemia-reperfusion injury, systemic inflammatory response syndrome, multiple organ failure, and permanent neurological damage (16,17). In this process, changes in hematological parameters can reveal both the severity of the damage and the body's potential for self-repair. In particular, changes in the activities of leukocyte subpopulations can provide valuable information about the severity of the systemic inflammatory response and patient prognosis (18,19).

The aim of this study is to evaluate the survival rates of patients who underwent CPR at the scene and the related demographic, clinical, and laboratory factors. The findings obtained may contribute to the identification of practical and effective biomarkers that can be used in early prognostic assessment after CPR.

Materials and Methods

This retrospective cohort study included 620 patients who underwent CPR between January 2020 and December 2022. The study was conducted with the approval of the İstanbul Medipol University Non-Interventional Clinical Research Ethics Committee (desicion number: E-10840098-772.02-3157, date: 01.07.2021).

Demographic characteristics (age, gender), clinical findings (level of consciousness, respiratory status, pulse rate), laboratory values, and time variables (command response time, station response time, transport time, intervention time, time to hospital arrival) were recorded.

Level of consciousness was assessed in three categories: conscious, confused, and unresponsive. Respiratory status was analysed in seven categories: normal, rapid, superficial, irregular, shortness of breath, none. Pulse status was assessed in four categories: normal, arrhythmic, thready and no pulse.

Laboratory values were assessed using venous and arterial blood samples taken upon patient admission. Parameters: white blood cell (WBC, 10^3/µL), red blood cell (RBC, 10_2/µL), hemoglobin (HGB, g/dL), hematocrit (HCT, %), platelet (PLT, 10^3/µL), neutrophil (NEUT, 10^3/µL), lymphocyte (LYMPH, 10^3/µL), monocyte (MONO, 10^3/µL), eosinophil (EO, 10^3/µL), basophil (BAS, 10^3/µL), creatinine (mg/dL), aspartate aminotransferase (AST, U/L), alanine aminotransferase (ALT, U/L), C-reactive protein (CRP, mg/L), and blood gas values (pH, PO_2 in mmHg, PCO_2 in mmHg) were included. Derived parameters such as the NLR and PLR were also calculated.

Successful CPR is defined as the restoration of spontaneous circulation and the recovery of sustained vital functions in a patient who has experienced cardiac arrest or respiratory arrest.

Long-term success: patients who survived for more than 24 hours were considered successful. Complete success: defined as 30-day survival and good neurological status (20).

Statistical Analysis

SPSS (Statistical Package for the Social Sciences) version 25.0 software was used for data analysis. Continuous variables are expressed as arithmetic distributed data, while categorical variables are expressed as frequency and percentage distributions.

The Shapiro-Wilk test was used to assess the normal distribution of the data. The Student's t-test was used to compare the quantitative data of two groups showing normal distribution. Pearson's chi-square analysis was used to compare frequencies. The relationship between survival and continuous variables was evaluated using Pearson correlation analysis.

Logistic regression analysis was used to determine the factors affecting survival. For the combined predictive model, pH and PO_2 values were first evaluated as continuous variables in Univariate logistic regression. Subsequently, optimal cut-off values for pH (≤6.817) and PO_2 (≤45.5 mmHg) were determined using receiver operating characteristic (ROC) curve analysis with Youden's index (sensitivity + specificity - 1) to maximize both sensitivity and specificity. These dichotomized variables, along with other significant hematological parameters, were then entered into a multivariate logistic regression model using the enter method. The predicted probabilities from the final model were used to construct a ROC curve, and the optimal cut-off probability (0.700) for mortality prediction was determined using Youden's index. The level of statistical significance was set at $p<0.05$.

Table 1. Demographic characteristics of patients

Gender, n (%)	
Female	220 (35.48)
Male	400 (64.52)
Region, n (%)	
Inland	591 (95.32)
External	29 (4.68)
Place of residence, n (%)	
Urban	611 (98.5)
Rural	9 (1.45)
Age, years, median (min-max)	67 (54-72)

Results

The median age of the 620 patients included in the study was 67 years (IQR: 54-79), and 64.52% (n=400) were male. 95.32% (n=591) of the calls received were made from within the region. 98.55% (n=611) of the patients lived in urban areas. Fully successful CPR was achieved in 7.90% (n=49) of these patients (Tables 1, 2).

12.10% of patients (n=75) had respiratory disease, 1.77% (n=11) had trauma due to a fall, 0.65% (n=4) had heart disease, 0.65% (n=4) had oncological disease, and 0.48% (n=3) had internal disease.

pH and PO_2 values were very low in deceased patients (p<0.001). PCO_2 values were similar in surviving and deceased patients (p=0.186). WBC, PLT, PLT/LYMPH ratio, BAS, MCV, and PCT were significantly lower in deceased patients (MCV: p=0.003, others: p<0.001). HGB, PLT/NEUT ratio, EO, MCH, MCHC, and PDW values were significantly higher in deceased patients (PLT/NEUT: p=0.038, other values: p<0.001) (Table 3).

In Univariate logistic regression analysis, the parameters with a significant effect on mortality were pH, pH ≤6.817 group variable, PO_2, and PO_2 ≤45.5 group variable, RBC, HGB, HCT, PLT, NEUT, LYMPH, PLT/LYMPH, NEUT/LYMPH, EO, BAS, MCV, MCH, MCHC, and PDW (Table 3).

In the multivariate logistic regression analysis, the mortality risk was 37.39 times higher in the pH ≤6.817 group (reference group: pH >6.817), [95% confidence interval (CI): 14.48-96.52, p=0.001].

For every 1-unit increase in values, the mortality risk is increased by 1% for PO2, 0.5% for PLT, 17% for NEUT/LYMPH, 82% for BASO, and 1% for MCHC (Table 3).

At a cut-off probability of 0.700 for the combined predictive model (pH + PO_2), the sensitivity of the predicted probabilities was 0.978, specificity was 0.831, positive predictive value was 0.976, and negative predictive value was 0.842 (Table 4).

The mortality prediction model, with an area under the curve (AUC) value of 0.923 (95% CI: 0.875-0.971), demonstrated significantly better performance than the pH and the PO_2 (DeLong test: p=0.017 for both) (Figure 1).

Table 2. Clinical characteristics of patients

Triage code, n (%)	
Yellow	15 (2.42)
Red	561 (90.48)
Black	44 (7.1)
Level of consciousness, n (%)	
Conscious	32 (5.17)
Confused	13 (2)
Unconscious	575 (92.7)
Pupils, n (%)	
Normal	66 (10.63)
Anisocoric	7 (1.08)
Fixed dilated	219 (35.52)
Isochoric	2 (0.36)
Myotic	18 (2.88)
Mydriatic	21 (3.41)
No response	287 (46.31)
Respiratory, n (%)	
Normal	18 (2.86)
Rapid	1 (0.19)
Superficial	7 (1.14)
Irregular	14 (2.29)
Shortness of breath	14 (2.29)
None	566 (91.24)
Skin, n (%)	
Normal	55 (8.91)
Cyanotic	321 (51.87)
Pale	217 (34.94)
Sweaty	17 (2.67)
Hyperemic	2 (0.36)
Icteric	2 (0.36)
Dry	6 (0.89)
Pulse, n (%)	
Normal	40 (6.53)
Arrhythmic	23 (3.73)
Filamentous	19 (2.99)
Not in use	538 (86.75)

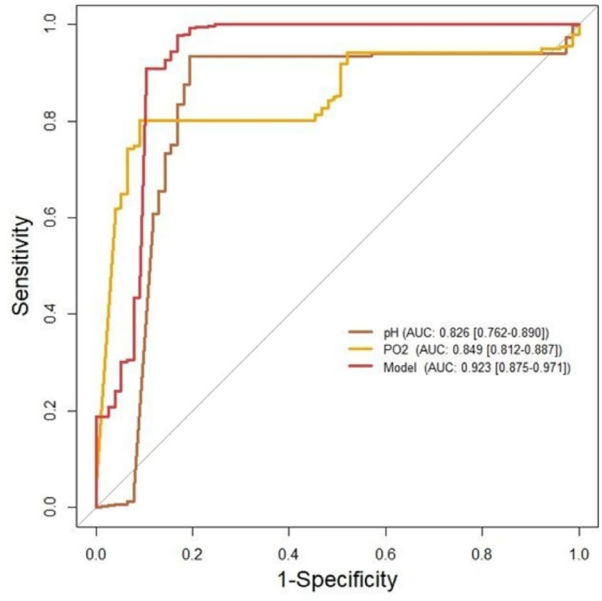


Figure 1. Receiver operating characteristic curves for mortality prediction using pH, PO_2, and the combined model (pH + PO_2)
AUC: Area under the curve

Table 3. Results of logistic regression analysis for mortality prediction using laboratory parameters						
	Univariate logistic regression analysis			Multivariate logistic regression analysis		
	Odds ratios	95% CI	p value	Odds ratios	95% CI	p value
pH	13.75	4.46-42.36	**<0.001**	-	-	-
pH ≤6.817	58.21	30.16-112.35	**<0.001**	37.39	14.48-96.52	**<0.001**
PCO_2	1.00	1.00-1.00	0.809	-	-	-
PO_2	0.990	0.986-0.993	**<0.001**	0.990	0.984-0.996	**0.002**
PO_2 ≤45.5	40.28	18.01-90.09	**<0.001**	-	-	-
WBC	1.00	0.99-1.00	0.369	-	-	-
RBC	1.64	1.29-2.09	**<0.001**	-	-	-
HGB	1.02	1.01-1.02	**<0.001**	-	-	-
HCT	1.07	1.04-1.10	**<0.001**	-	-	-
PLT	0.98	0.98-0.99	**<0.001**	0.99	0.99-1	**0.034**
NEUT	0.97	0.96-0.98	**<0.001**	-	-	-
LYMPH	1.03	1.02-1.05	**<0.001**	-	-	-
PLT/NEUT	0.96	0.91-1.01	0.116	-	-	-
PLT/LYMPH	0.91	0.89-0.93	**<0.001**	-	-	-
NEUT/LYMPH	0.75	0.70-0.81	**<0.001**	0.83	0.72-0.94	**0.004**
MONO	1.00	0.98-1.03	0.855	-	-	-
EO	1.60	1.28-1.99	**<0.001**	-	-	-
BAS	0.05	0.02-0.13	**<0.001**	0.18	0.06-0.55	**0.002**
MCV	1.05	1.01-1.08	**0.009**	-	-	-
MCH	1.39	1.24-1.56	**<0.001**	1.16	0.97-1.39	0.098
MCHC	1.01	1.00-1.01	**<0.001**	1.01	1.01-1.01	**<0.001**
MPV	1.13	0.88-1.44	0.340	-	-	-
PCT	0.94	0.74-1.20	0.632	-	-	-
PDW	2.39	1.52-3.75	<0.001	1.52	0.96-2.39	0.074

PLT: Platelet, NEUT: Neutrophil, LYMPH:Lymphocyte, WBC: White blood cell, RBC: Red blood cell, HGB: Hemoglobin, HCT: Hematocrit, MONO: Monocyte, EO: Eosinophil, BAS: Basophil, MCV: Mean corpuscular volume, MCH: Mean corpuscular hemoglobin, MCHC: Mean corpuscular hemoglobin concentration, MPV: Mean platelet volume, PCT: Plateletcrit, PDW: Platelet distribution width

Table 4. Diagnostic performance of pH, PO_2, and the combined model (pH + PO_2)			
	Combined model (pH + PO_2)	pH	PO_2
AUC (95% CI)	0.923 (0.875-0.971)	0.826 (0.762-0.890)	0.849 (0.812-0.887)
Cut-off point	0.700	6.817	45,500
Sensitivity	0.978	0.934	0.801
Specificity	0.831	0.805	0.909
Positive predictive value	0.976	0.971	0.984
Negative predictive value	0.842	0.633	0

AUC: Area under the curve, CI: Confidence interval

Discussion

This study aimed to investigate the effect of sociodemographic, clinical, and laboratory parameters on mortality in patients undergoing CPR during cardiac arrest monitoring. A retrospective analysis of 620 patients revealed the role of arterial blood gas parameters and hematological biomarkers in predicting mortality risk.

A study published in 2024 observed that 67.5% of out-of-hospital cardiac arrest cases were in patients over 65 years of age (21).

The median age of the patients included in the study was 67 years (IQR: 54-79), and 64.5% were male, supporting that cardiac arrest is more common in older individuals and males (22). This finding is consistent with the literature showing that cardiovascular risk factors increase with age and are more prevalent in males.

98.5% of patients lived in urban areas indicates that intervention at the scene was faster. However, our study found no significant difference in positively affecting survival rates after CPR. This indicates that the effectiveness of CPR should not be evaluated based solely on geographical location.

The successful resuscitation rate is only 7.9%, and a study reported in the literature indicates that out-of-hospital cardiac arrest survival rates are 9% in Europe, 6% in North America, 11% in Australia, and 2% in Asia (23). Our findings are consistent with international values.

Recent evidence from the 2024 update of the Utstein Out-of-Hospital Cardiac Arrest Reporting Template provides an important context for interpreting survival rates (24). The updated guidelines emphasise the importance of standardised reporting and risk adjustment for key Utstein factors such as age, gender, location of arrest, and bystander status. When evaluated against these current standards, our observed survival rate reflects the complex interaction of multiple prognostic factors.

A key finding from recent research shows that cardiac arrest survival rates deteriorated significantly during the coronavirus disease 2019 pandemic, with survival rates falling significantly in 2020 and remaining below pre-pandemic levels (25). This temporal context is particularly relevant to our study period (2020-2022) and suggests that the survival rates we observed may reflect both traditional prognostic factors and pandemic-related healthcare system challenges.

Although there were variations between countries, a generally low survival rate was observed. Possible reasons include delayed intervention, failure to recognise cardiac arrest, or serious comorbidities before cardiac arrest. Since 95% of calls were due to medical reasons, with 12% of these being respiratory system diseases, it indicates that acute respiratory decompensation plays a significant role in the development of cardiac arrest.

In our study, analyses of laboratory data revealed that arterial blood gas parameters, particularly pH and PO_2 values, were the strongest predictors of mortality. In Univariate and multivariate logistic regression analyses, the mortality risk was 37.4 times higher in patients with pH ≤6.817. This finding indicates that metabolic acidosis developing after arrest is incompatible with life. von Auenmueller et al. (26) specifically investigated the value of arterial blood gas parameters for predicting mortality in out-of-hospital cardiac arrest survivors, and reported that pH and lactate were the most relevant parameters because they were strongly and independently associated with mortality.

Hypoxaemia was found to be similarly an independent predictor of mortality, with the mortality rate increasing dramatically below PO_2 ≤45.5 mmHg. These findings support the critical role of oxygenation in resuscitation success and the need to integrate early arterial blood gas measurements into clinical decision support systems (27).

Hematological parameters also yielded noteworthy results. PLT counts were significantly higher in survivors, and each unit increase in PLT levels was associated with a 0.5% reduction in mortality risk. This demonstrates that microvascular dysfunction and systemic inflammation following cardiac arrest can be monitored via hematological markers (28).

In particular, the PLT/LYMPH ratio and NEUT/LYMPH ratio provide information about the nature of the inflammatory response. The increase in mortality with a decrease in the NEUT/LYMPH ratio suggests that suppression of the immune response after cardiac arrest may be an indicator of poor prognosis. This finding is consistent with studies showing that immune functions play an important role in the early period after cardiac arrest (29).

The inverse relationship between BAS and mortality, and the significant difference in EO values, are also noteworthy. These findings may reflect the immunomodulatory effects of the granulocyte series and suggest that some hematological parameters may represent the disruption of immune balance after arrest. High erythrocyte indices such as MCH and MCHC likely reflect the compensatory response to hypoxic stress and oxygen-carrying capacity after arrest, and this is consistent with the literature (30).

In our study, the predictive power of the model was also evaluated using ROC analysis. The AUC value of the model including the pH and PO_2 parameters was found to be 0.923 (95% CI: 0.875-0.971). This high AUC value indicates that the model has a strong discriminatory power in predicting mortality.

Specifically, when using a threshold value of 0.700, the sensitivity and specificity of the model were calculated as 97.8% and 83.1%, respectively. This performance may provide significant advantages in clinical practice for early prognosis determination (31).

Study Limitations

The main limitations of this study are its retrospective nature and the absence of certain clinical variables (CPR duration, witness to cardiac arrest).

Conclusion

In conclusion, this study demonstrated that arterial blood gas findings (particularly ≤6.817 and PO_2 ≤45.5 mmHg) and certain hematological markers (PLT, NEUT/LYMPH ratio, BAS) have high diagnostic value in predicting mortality after cardiac arrest. These findings are promising as they can be used in early triage and clinical decision-making processes. Prospective and multicentre studies could strengthen the integration of these parameters into clinical practice.

Ethics

Ethics Committee Approval: The study was conducted with the approval of the İstanbul Medipol University Non-Interventional Clinical Research Ethics Committee (desicion number: E-10840098-772.02-3157, date: 01.07.2021).

Informed Consent: This retrospective study.

Footnotes

Author Contributions

Surgical and Medical Practices: K.A.T., Concept: M.K., Design: M.K., Data Collection or Processing: M.K., Analysis or Interpretation: G.A., Literature Search: A.Y.G., Writing: G.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. McNally B, Robb R, Mehta M, Vellano K, Valderrama AL, Yoon PW, et al. Out-of-Hospital Cardiac Arrest Surveillance — Cardiac Arrest Registry to Enhance Survival (CARES), United States, 1 October 2005–31 December 2010. MMWR Surveill Summ. 2011;60:1-19.
2. Perkins GD, Graesner JT, Semeraro F, Olasveengen T, Soar J, Lott C, et al. European Resuscitation Council Guidelines 2021: summary. Resuscitation. 2021;161:1-60.
3. Berdowski J, Berg RA, Tijssen JG, Koster RW. Global incidence and survival rates of out-of-hospital cardiac arrest: a systematic review of 67 prospective studies. Resuscitation. 2010;81:1479-87.
4. Gräsner JT, Wnent J, Herlitz J, Perkins GD, Lefering R, Tjelmeland I, et al. Survival after out-of-hospital cardiac arrest in Europe – results of the EuReCa TWO study. Resuscitation. 2020;148:218-26.
5. Andersen LW, Holmberg MJ, Berg KM, Donnino MW, Granfeldt A. In-hospital cardiac arrest: a review. JAMA. 2019;321:1200-10.
6. Hawkes C, Booth S, Ji C, Brace-McDonnell SJ, Cooke MW, Deakin CD, et al. Epidemiology and outcomes of out-of-hospital cardiac arrest in England. Resuscitation. 2017;110:133-40.
7. Girotra S, Nallamothu BK, Spertus JA, Li Y, Krumholz HM, Chan PS, et al. Trends in survival after in-hospital cardiac arrest. N Engl J Med. 2012;367:1912-20.
8. Şener A, Pekdemir M, İslam MM, Yılmaz S, Çorbacıoğlu ŞK, Çevik Y, et al. Prospective, multicenter, Turkish out-of-hospital cardiac arrest study: TROHCA. Turk J Emerg Med. 2024:24;133-44.
9. Sasson C, Rogers MA, Dahl J, Kellermann AL. Determinants of survival after out-of-hospital cardiac arrest: a systematic review and meta-analysis. Circ Cardiovasc Qual Outcomes. 2010;3:63-81.
10. Martinell L, Nielsen N, Herlitz J, Karlsson T, Horn J, Wise MP, et al. Early predictors of poor outcomes after out-of-hospital cardiac arrest. Crit Care. 2017;21:96.
11. Dell'Anna AM, Sandroni C, Lamanna I, Belloni I, Donadello K, Creteur J, et al. Prognostic significance of blood lactate concentrations after cardiac arrest: a retrospective study. Ann Intensive Care. 2017;7:101.
12. Düring J, Dankiewicz J, Cronberg T, Hassager C, Hovdenes J, Kjaergaard J, et al. Post-cardiac arrest lactate, lactate clearance and outcomes: a post-hoc analysis of the TTM-trial. Acta Anaesthesiol Scand. 2018;62:1436-42.
13. Chang CJ, Liou TH, Tsai WT, Huang SW, Liu CW, Huang ST, et al. Clinical and hematological predictors of return of spontaneous circulation in patients with out-of-hospital cardiac arrest. J Acute Med. 2020;10:51-9.
14. Balta S, Celik T, Mikhailidis DP, Ozturk C, Demirkol S, Aparci M, et al. The relationship between atherosclerosis and the neutrophil-lymphocyte ratio. Clin Appl Thromb Haemost. 2016;22:405-11.
15. Li H, Zhou Y, Ma Y, Han S, Li L, Wang D, et al. Prognostic value of the platelet-lymphocyte ratio in acute coronary syndrome: a systematic review and meta-analysis. Cardiol Pol. 2017;75:666-73.
16. Adrie C, Adib-Conquy M, Laurent I, Cariou A, Dhainaut JF, Spaulding C, et al. Successful cardiopulmonary resuscitation after cardiac arrest as a "sepsis-like" syndrome. Circulation. 2002;106:562-8.
17. Nolan JP, Neumar RW, Adrie C, Aibiki M, Berg RA, Böttiger BW, et al. Post-cardiac arrest syndrome: epidemiology, pathophysiology, treatment, and prognosis. Resuscitation. 2008;79:350-79.
18. Peberdy MA, Andersen LW, Abbate A, Thacker LR, Gaieski D, Abella B, et al. Inflammatory markers following out-of-hospital cardiac arrest resuscitation: a prospective multicentre observational study. Resuscitation. 2016;103:117-24.
19. Vaahersalo J, Skrifvars MB, Pulkki K, Stridsberg M, Røsjø H, Hovilehto S, et al. Interleukin-6 at admission is associated with post-resuscitation organ dysfunction and predicts long-term neurological outcome after out-of-hospital ventricular fibrillation. Resuscitation. 2014;85:1573-9.
20. Perkins GD, Graesner JT, Semeraro F, Olasveengen T, Soar J, Lott C, et al. European Resuscitation Council Guidelines 2021: executive summary. Resuscitation. 2021;161:1-60.
21. Chen CY, Fan CY, Chen IC, Hsu CY, Su YC, Huang CC, et al. The interaction of gender and age on outcomes in out-of-hospital cardiac arrest cases treated with emergency medical services: a 5-year multicentre retrospective analysis, Resuscitation Plus. 2024;17:100552.
22. Nichol G, Thomas E, Callaway CW, Hedges J, Powell JL, Aufderheide TP, et al. Regional variations in out-of-hospital cardiac arrest incidence and outcomes. JAMA. 2008;300:1423-31.

23. Mathiesen WT, Bjørshol CA, Kvaløy JT, Søreide E. Effects of modifiable prehospital factors on survival after out-of-hospital cardiac arrest in rural versus urban areas. Crit Care. 2018;22:99.

24. Bray JE, Grasner JT, Nolan JP, Iwami T, Ong MEH, Finn J, et al. Cardiac arrest and cardiopulmonary resuscitation outcome reports: 2024 update of the Utstein out-of-hospital cardiac arrest registry template. Circulation. 2024;149:1331-67.

25. American Heart Association. Cardiac arrest survival improved since COVID-19 pandemic waned, still lower than prior years. American Heart Association News. Available from: https://newsroom.heart.org/news/cardiac-arrest-survival-improved-since-covid-19-pandemic-waned-still-lower-than-prior-years

26. von Auenmueller KI, Christ M, Sasko BM, Trappe HJ. The value of arterial blood gas parameters for prediction of mortality in survivors of out-of-hospital cardiac arrest. J Emerg Trauma Shock. 2017;10:134-9.

27. Hong SI, Kim JS, Kim YJ, Lee SW, Lee SU, Jo YH, et al. Dynamic changes in arterial blood gas during cardiopulmonary resuscitation in out-of-hospital cardiac arrest. Sci Rep. 2021;11:23165.

28. Taha Sert E, Kokulu K, Mutlu H, Gül M, Uslu Y. Performance of the systemic immune-inflammation index in predicting the likelihood of discharge in out-of-hospital cardiac arrest. Resusc Plus. 2023;14:100382.

29. Salciccioli JD, Marshall DC, Pimentel MA, Santos MD, Pollard T, Celi LA, et al. The relationship between the neutrophil-lymphocyte ratio and mortality in critically ill patients: an observational cohort study. Crit Care. 2015;19:13.

30. Erol MK, Kankılıc N, Kaya F, Çelik MM, Gürsoy MO, Özkan B, et al. The relationship between hematological parameters and mortality in cardiovascular patients with post-cardiac arrest syndrome. Cureus. 2019;11:6478.

31. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparison of areas under two or more correlated ROC curves: a nonparametric approach. Biometrics. 1988;44:837-45.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.84770
Eurasian J Emerg Med. 2026;25: 45-52

The Effect of Delirium Knowledge Level Among Intensive Care Patients' Relatives on Family Care Satisfaction: A Cross-Sectional Study

Tuğba Menekli

Malatya Turgut Özal University Faculty of Health Sciences, Department of Nursing, Malatya, Türkiye

Abstract

Aim: Delirium affects 20-80% of intensive care unit (ICU) patients and significantly impacts family members' psychological well-being and care satisfaction. However, the relationship between family members' delirium knowledge and their satisfaction with ICU care remains understudied. This study investigated the association between ICU patients' relatives' delirium knowledge levels and their care satisfaction, while identifying key predictors of family satisfaction.

Materials and Methods: This cross-sectional study was conducted between January and April 2024 at Turgut Özal Medical Center, Malatya, Türkiye. The study included 305 family members of ICU patients hospitalized for ≥48 hours. Participants were recruited using convenience sampling from four ICUs (internal medicine, coronary, neurology, and medical oncology). Data collection instruments included: Caregiver Characteristics Form, Caregiver Intensive Care Delirium Knowledge Questionnaire-Turkish Version (CIDKQ-T), and Family Satisfaction in the Intensive Care Unit scale (FS-ICU-24). Statistical analyses included descriptive statistics, Pearson correlation coefficient/Spearman's rank correlation, and multiple linear regression.

Results: Participants demonstrated moderate delirium knowledge (CIDKQ-T mean: 11.62±2.54, range: 0-21) and moderate ICU care satisfaction (FS-ICU-24 mean: 52.14±14.48, range: 0-100). Strong positive correlations were found between delirium knowledge and all FS-ICU-24 dimensions: total satisfaction ($r=0.60$, $p<0.001$), care satisfaction ($r=0.45$, $p<0.001$), decision-making satisfaction ($r=0.53$, $p<0.001$), and information satisfaction ($r=0.48$, $p<0.001$). Multiple regression analysis ($R^2=0.36$, $F(7.142) = 12.85$, $p<0.001$) identified delirium knowledge as the strongest predictor of family satisfaction ($\beta=0.42$, $p<0.001$), followed by university education ($\beta=0.31$, $p=0.002$), presence of neurological disease ($\beta=0.28$, $p=0.004$), and female gender ($\beta=0.25$, $p=0.008$). Longer hospitalization (10-13 days) negatively affected satisfaction ($\beta=-0.22$, $p=0.018$).

Conclusion: Family members' delirium knowledge significantly predicts their ICU care satisfaction, explaining 36% of the variance in satisfaction scores. These findings support implementing structured delirium education programs for families as part of family-centered ICU care protocols.

Keywords: Delirium knowledge, intensive care unit, family satisfaction, patient relatives

Introduction

Intensive care units (ICUs) constitute specialized clinical environments dedicated to the continuous monitoring and management of critically ill patients utilizing advanced technology and a multidisciplinary team approach. These units are indispensable for patients necessitating intensive medical and nursing support (1,2). Patients within ICUs encounter substantial physiological and psychological stressors, which markedly elevate their susceptibility to delirium. Delirium is a neurocognitive disorder characterized by acute onset, fluctuating course, and impairments in attention, awareness, and cognition (3,4). It ranks among the most prevalent complications in ICUs, with reported incidence rates varying from 20% to 80%, and even higher frequencies observed in patients undergoing mechanical ventilation (3-5).

Corresponding Author: Tuğba Menekli MD, Malatya Turgut Özal University Faculty of Health Sciences, Department of Nursing, Malatya, Türkiye
E-mail: tugba.menekli@ozal.edu.tr **ORCID ID:** orcid.org/0000-0002-8291-8274
Cite this article as: Menekli T. The effect of delirium knowledge level among intensive care patients' relatives on family care satisfaction: a cross-sectional study. Eurasian J Emerg Med. 2026;25: 45-52.

Received: 31.08.2025
Accepted: 29.09.2025
Epub: 23.10.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

The detrimental clinical outcomes associated with delirium are extensively documented; it is known to prolong hospitalization, elevate mortality risk, and contribute to long-term cognitive impairment (6,7). Affected patients often require extended periods of mechanical ventilation, experience heightened rates of complications, and demonstrate reduced functional independence following discharge. Significantly, the repercussions of delirium extend beyond the patient, profoundly impacting their family members (2,3).

Family members witnessing their loved ones exhibiting symptoms such as altered consciousness, agitation, disorientation, and uncharacteristic behavioral changes frequently endure intense psychological distress and uncertainty (8,9). These experiences can precipitate anxiety, depressive symptoms, and post-traumatic stress among relatives (10). Particularly, individuals lacking adequate understanding of delirium may misinterpret these manifestations, leading to feelings of helplessness and apprehension. This can subsequently impair communication and erode trust between families and healthcare providers, adversely affecting overall satisfaction with the ICU experience (11,12).

Emerging research underscores a positive correlation between the level of knowledge families possess about delirium and their reported satisfaction with care received in the ICU (13,14). When relatives comprehend the etiology and clinical features of delirium, they are better equipped to participate meaningfully in the care process and engage in more effective communication with healthcare staff (10,11). Educational and awareness-raising initiatives have demonstrated efficacy in reducing family anxiety and enhancing their capacity to navigate the ICU experience (2,15). Consequently, providing accurate and comprehensive information concerning delirium is paramount—not solely for improving patient outcomes but also for safeguarding the psychosocial well-being of families and their satisfaction with healthcare services (6,16).

A review of contemporary literature indicates a relative paucity of studies comprehensively investigating the relationship between family members' knowledge of delirium and their satisfaction with ICU care (17,18). Much of the existing research focuses on general ICU experiences, often lacking a detailed examination of the unique psychosocial burden imposed by delirium on families (8,18). Given its distinct clinical trajectory and associated uncertainties, delirium presents a specific emotional and psychological challenge for family caregivers, differentiating it from other ICU complications.

In light of these considerations, this study concurrently assesses the knowledge levels of ICU patients' family members regarding delirium and their satisfaction with the care provided. The outcomes are anticipated to furnish a foundation for developing targeted educational strategies, foster active family involvement in delirium management, and reinforce patient- and family-centered care approaches. Furthermore, the study explores demographic and clinical factors influencing this relationship and identifies significant predictors of family care satisfaction.

Materials and Methods

Study Design

This investigation was conceived as a cross-sectional study, conducted over a four-month period from January to April 2024 at the Turgut Özal Medical Center, affiliated with Malatya Turgut Özal University, in Malatya, Türkiye.

Sample and Setting

The study population comprised family members of patients hospitalized in the internal medicine, coronary, neurology, and medical oncology ICUs of the aforementioned medical center. The final sample included 305 family members of patients who had been admitted to the ICU for a minimum of 48 hours. Sample size determination was performed using G*Power 3.1 software, specifying an effect size of 0.50, a statistical power of 95%, and an alpha significance level of 0.05. This calculation yielded a minimum required sample size of 305 participants, consistent with parameters derived from prior correlation studies examining analogous relationships (9,16).

Inclusion Criteria:

- Age 18 years or older
- Primary caregiver designation for an ICU patient hospitalized for ≥48 hours
- Minimum of three separate visits to the patient during the ICU stay
- Demonstrated ability to comprehend and complete the study questionnaires
- Provision of voluntary informed consent to participate

Exclusion Criteria:

- Presence of visual, auditory, or cognitive impairments that would preclude completion of the questionnaires
- History of a prior psychiatric diagnosis potentially influencing responses
- Inability or unwillingness to provide informed consent

Data Collection Tools

Caregiver Characteristics Form: This instrument was developed by the research team following an extensive review of relevant

literature. It was designed to capture participants' socio-demographic details (e.g., age, gender, educational background) and attributes related to their caregiving role (e.g., relationship to the patient, prior caregiving experience).

Caregiver Intensive Care Unit Delirium Knowledge Questionnaire-Turkish Version (CIDKQ-T): The CIDKQ was developed by Krewulak et al. (19) to quantify caregivers' knowledge concerning delirium in ICU patients. The Turkish adaptation and validation of this tool was subsequently performed by Erbay Dallı and Kelebek Girgin (20), who reported a test-retest reliability coefficient of 0.81. The questionnaire encompasses 21 items distributed across three subscales: risk factors (10 items), management practices (6 items), and symptoms (5 items). Respondents select from "Yes," "No," or "I don't know" options. Correct responses are assigned 1 point, while incorrect or "I don't know" responses receive 0 points. Total scores range from 0 to 21, with higher scores indicative of greater knowledge about delirium.

Family Satisfaction in the Intensive Care Unit (FS-ICU-24): FS-ICU-24 is a validated instrument initially developed by Heyland and Tranmer (21), and later refined by Wall et al. (22), designed to systematically evaluate family members' satisfaction with care provided in the ICU. The validity and reliability of the Turkish version were established by Tastan et al. (23). The scale comprises 24 items organized into two primary domains: satisfaction with care (14 items) and satisfaction with decision-making (10 items). Responses are recorded on a 5-point Likert scale. Higher scores correspond to higher levels of satisfaction. In the current study, Cronbach's alpha values were calculated as 0.96 for the overall scale, 0.96 for the care subscale, 0.82 for the decision-making subscale, and 0.92 for the information subscale.

Data Collection Process

Data acquisition was conducted through face-to-face interviews administered by the researchers. Participants completed the Caregiver Characteristics Form, the CIDKQ-T, and the FS-ICU-24 under researcher supervision to ensure clarity and comprehensiveness of responses.

Statistical Analysis

All statistical analyses were performed using IBM SPSS Statistics version 28.0. Descriptive measures, including frequencies, percentages, means, and standard deviations, were calculated to summarize the study variables. The Kolmogorov-Smirnov and Shapiro-Wilk tests were used to assess data normality, while relationships between variables were examined through Pearson correlation coefficient or Spearman's rank correlation analyses, as appropriate. Multiple linear regression analysis was subsequently conducted to identify the key predictors of family satisfaction with ICUs. Statistical significance was defined as a p-value less than 0.05 for all analyses.

Ethical Considerations

This study received approval from the Non-Interventional Clinical Ethics Committee of İnönü University (desicion number: 2024-6/1, date: 14.01.2024). Formal institutional permission was also obtained from the administration of the Turgut Özal Medical Center. The research was conducted in strict adherence to the principles outlined in the Declaration of Helsinki. All participants were comprehensively informed about the study's purpose, scope, and procedures, and their participation was entirely voluntary. Written informed consent was obtained from each individual prior to enrolment. The confidentiality of all collected information was rigorously maintained, and data were used exclusively for scientific purposes.

Results

Among the 305 participating family members, 190 (62.3%) were female. The mean age of participants was 48.98±7.05 years. The largest proportion of participants (n=135, 44.3%) fell within the 31-41 age range. The majority reported being married (64.1%), having attained a high school education (45.3%), and having income levels that approximately matched their expenses (43.0%). In terms of relationship to the patient, children constituted the largest group (37.7%). A significant majority of participants (68.9%) reported no prior experience with patient care. Regarding the primary diagnoses of the hospitalized patients, cardiovascular diseases were most prevalent (32.1%), while oncological diseases were least common (13.8%). Analysis of hospitalization duration indicated that 40% of patients remained in the ICU for a period of 10-13 days (Table 1).

On the FS-ICU-24, the mean total satisfaction score was 52.14±14.48. Examination of the subscale scores revealed means of 54.78±19.0 for the care subscale, 49.07±18.48 for the decision-making subscale, and 16.91±16.13 for the Information subscale. Performance on the CIDKQ-T yielded a mean total score of 11.62±2.54. Subscale analysis showed mean scores of 6.62±1.5 for risk factors, 3.56±0.5 for management practices, and 1.94 ± 0.2 for symptoms (Table 2).

A robust positive correlation was identified between the total score on the FS-ICU-24 and the total score on the CIDKQ-T ($r=0.60$, $p<0.001$). Significant positive correlations were also observed between the FS-ICU-24 care subscale and the CIDKQ-T risk factors subscale ($r=0.461$, $p<0.001$), the CIDKQ-T practices subscale ($r=0.484$, $p<0.001$), and the CIDKQ-T symptoms subscale ($r=0.471$, $p<0.001$).

Furthermore, significant positive relationships were found between the FS-ICU-24 decision-making subscale and the CIDKQ-T risk factors subscale (r=0.401, p=0.001), the CIDKQ-T practices subscale (r=0.493, p=0.001), and the CIDKQ-T symptoms subscale (r=0.516, p=0.001).

Table 1. Socio-demographic characteristics of the participants

Variables	Number	Percentage (%)
Gender		
Male	115	37.7
Female	190	62.3
Age groups		
20-30	90	29.5
31-41	135	44.3
42-52	80	26.2
Marital status		
Married	157	64.1
Single	88	35.9
Educational level		
Elementary school graduate	98	32.1
High school graduate	138	45.3
University graduate	69	22.6
Level of income		
Income is lower than expenses	93	30.5
Income equals expenses	131	43.0
The income of the organization is higher than its expenses	81	26.5
Degree of closeness with the patient		
Wife	50	16.4
Parent	68	22.3
Child	115	37.7
Sibling	72	23.6
Previous patient care experience		
Yes	95	31.1
No	210	68.9
Diagnosis of patients in intensive care units		
Cardiovascular diseases	98	32.1
Respiratory system diseases	69	22.6
Neurological diseases	46	15.1
Oncological diseases	42	13.8
Infections	50	16.4
Length of hospitalization		
3-6 days	56	18.4
6-9 days	71	23.2
10-13 days	122	40.0
14 or more days	56	18.4
Average age	48.98±26.5	

Similarly, significant positive correlations were detected between the FS-ICU-24 information subscale and the CIDKQ-T risk factors subscale (r=0.437, p<0.001), the CIDKQ-T practices subscale (r=0.526, p<0.001), and the CIDKQ-T symptoms subscale (r=0.501, p<0.001) (Table 3).

Multiple linear regression analysis identified several factors as statistically significant predictors of family care satisfaction among ICU patients' relatives [F (7,142)=12.85, p<0.001)]. The regression model explained 36% of the variance in the dependent variable (R^2=0.36, adjusted R^2=0.34). The most substantial predictors were identified as the level of knowledge about delirium (β=0.42, p<0.001; standardized β=0.38) and attainment of a university-level education (β=0.31, p=0.002; Standardized β=0.28). The presence of a neurological disease in the patient (β=0.28, p=0.004) and being female (β=0.25, p=0.008) also demonstrated significant positive relationships with satisfaction. Conversely, a patient's hospitalization duration of 10-13 days was associated with a negative effect on care satisfaction (β=-0.22, p=0.018). Demographic factors such as younger age (20-30 years age group; β=0.18, p=0.032) and being the child of the patient (β=0.15, p=0.045) exhibited weaker, yet statistically significant, positive effects. Multicollinearity diagnostics (VIF values all being <5) and the Durbin-Watson statistic (1.98) confirmed that the underlying assumptions of the regression model were not violated. These results underscore the critical influence of delirium knowledge and educational background on family care satisfaction (Table 4).

Discussion

This research investigated the knowledge levels concerning delirium among relatives of patients receiving treatment in ICU and explored its relationship with their satisfaction. The finding that 62% of participants were female aligns with existing literature, which consistently shows that caregiving

Table 2. Mean scores of the participants on the scales

Scale names	Mean ± SD	Min-max
Intensive care family satisfaction scale total score	52.14±14.48	0-100
Care subdimension	54.78±19.0	0-100
Decision making subdimension	49.07±18.48	0-100
Knowledge subdimension	16.91±16.13	0-100
Intensive care delirium knowledge level test for caregivers total score	11.62±2.54	0-21
Risk factors sub-dimension	6.62±1.5	0-10
Applications sub-dimension	3.56±0.5	0-6
Symptoms subscale	1.94±0.2	0-5

SD: Standard deviation

responsibilities disproportionately fall on women. Studies frequently indicate that women assume more active roles in patient care and are typically designated as primary caregivers (18,24,25). This observation reinforces the significant role gender plays in healthcare dynamics and supports the notion that women are more frequently engaged in the intensive care process. The concentration of participants (44.3%) within the 31 to 41-year age range suggests that adults in this age group are particularly likely to be involved in ICU caregiving. The overall mean age of 48.98±7.05 years indicates that the relatives of ICU patients are generally middle-aged. This implies that the demographic plays a crucial role as caregivers and that age is a factor influencing active participation in care processes (18,26). The high proportion of married individuals (64.1%) suggests that spousal and family support networks are instrumental in the caregiving process. Married individuals may have access to greater emotional and practical support, which constitutes an important resource for coping with stress (24,25). Additionally, the finding that 68.9% of participants had no prior patient care experience highlights potential knowledge gaps and uncertainties that could exacerbate anxiety during the care process. This underscores the vital importance of providing information and support specifically tailored to novice caregivers.

The predominance of children (37.7%) as the primary relatives involved indicates that the responsibility for caring for ICU patients often falls to the younger generation. This finding suggests a shifting dynamic where children are playing an increasingly important role in intensive care processes. Analysis of patient diagnoses revealed that cardiovascular diseases were most common (32.1%), reflecting the high prevalence of these conditions in ICUs and their frequent requirement for complex and prolonged care. In contrast, oncological diseases had the lowest representation (13.8%), suggesting that cancer patients may require or receive intensive care less frequently. This points to potentially different care dynamics and trajectories for oncology patients, who often have distinct treatment needs and health challenges (24-26).

Previous studies have generally reported that caregiver scores on delirium knowledge tests like the CIDKQ-T tend to be at or below average levels (27-30). This indicates that while caregivers may possess basic awareness, they often lack comprehensive knowledge about delirium. Studies utilizing the FS-ICU-24 scale typically report average total satisfaction scores ranging from moderate to high (15,28), suggesting that families are generally satisfied with ICU care. When examining subscales, the "care" subscale often receives the highest scores, implying that families are particularly satisfied with the direct medical and nursing care provided (27). Conversely, scores on the "information" and "decision-making" subscales are frequently lower, indicating that families may feel less informed about their relative's condition and less involved in care decisions (27). These patterns may vary based on healthcare system structures, cultural contexts, and methodological differences across studies conducted in different regions (26,27). Nonetheless, these consistent trends highlight

Table 3. Correlations between FS-ICU-24 and CIDKQ-T scales and subscales

	"FS-ICU-24 care subscale	FS-ICU-24 decision making subscale	FS-ICU-24 knowledge subscale	FS-ICU-24 total score
CIDKQ-T total score	r=0.450, p=0.000*	r=0.528, p=0.001*	r=0.477, p=0.000*	r=0.600, p=0.000*
CIDKQ-T risk factors subscale	r=0.461, p=0.000*	r=0.401, p=0.001*	r=0.437, p=0.000*	r=0.506, p=0.000*
CIDKQ-T applications subscale	r=0.484, p=0.000*	r=0.493, p=0.001*	r=0.526, p=0.000*	r=0.456, p=0.000*
CIDKQ-T symptoms subscale	r=0.471, p=0.000*	r=0.516, p=0.001*	r=0.501, p=0.000*	r=0.541, p=0.000*

*Spearman's correlation p<0.05. CIDKQ-T: Caregiver Intensive Care Unit Delirium Knowledge Questionnaire-Turkish version, FS-ICU-24: Family Satisfaction in the Intensive Care Unit

Table 4. Factors predicting family care satisfaction in relatives of intensive care unit patients

Independent variables	β	Stand. error	p-value	95% CI	Stand. β	VIF
Delirium knowledge level	0.42	0.08	<0.001	0.26-0.58	0.38	1.20
Education level (university)	0.31	0.07	0.002	0.17-0.45	0.28	1.15
Gender (female)	0.25	0.06	0.008	0.13-0.37	0.22	1.10
Age group (20-30 years)	0.18	0.05	0.032	0.03-0.33	0.16	1.08
Degree of closeness (child)	0.15	0.04	0.045	0.07- 0.23	0.14	1.05
Length of hospitalization (10-13)	-0.22	0.05	0.018	-0.32-0.12	-0.20	1.12
Presence of neurological disease	0.28	0.06	0.004	0.16-0.40	0.25	1.18

β: Beta, CI: Confidence interval, VIF: Variance inflation factor

the need to promote family-centered care practices, ensure adequate information provision and emotional support for families, and actively involve them in decision-making processes within ICUs to enhance overall satisfaction (10,25). Therefore, the current findings reinforce the necessity for continued efforts to improve family satisfaction in ICU settings.

The strong positive correlation identified between the total scores of the FS-ICU-24 and the CIDKQ-T ($r=0.60$, $p<0.001$) demonstrates that ICU patients' relatives' knowledge level is directly associated with their satisfaction with the care experience. This aligns with literature suggesting that adequate information about the ICU process positively influences family members' perceptions of care quality and facilitates their psychosocial adjustment (13,14). Conversely, insufficient information is known to exacerbate feelings of stress, anxiety, and uncertainty among relatives, thereby negatively impacting overall satisfaction levels (28).

The significant correlations between the FS-ICU-24 care subscale and the various CIDKQ-T subscales (risk factors, applications, symptoms) indicate that relatives' understanding of the critical illness process directly affects their perception of care quality. Studies consistently emphasize that families require comprehensive information to comprehend their relative's condition and that transparent information-sharing practices are strongly linked to care satisfaction (13,28). Specifically, providing sufficient details regarding disease management, treatment protocols, and symptom control enables family members to collaborate more effectively with the healthcare team (14,27).

The positive relationships between the FS-ICU-24 decision making subscale and the CIDKQ-T subscales suggest that as relatives' knowledge increases, so does their participation in decision-making processes. Literature confirms that involving family members in decisions is critical from both ethical and psychosocial perspectives (24,26). Active participation reduces feelings of uncertainty and strengthens their sense of responsibility and engagement in their loved one's treatment journey (10,25,26). Moreover, being well-informed allows families to make more considered choices regarding invasive interventions, palliative care options, and end-of-life preferences (28).

The significant correlations between the FS-ICU-24 information subscale and the CIDKQ-T subscales underscore that the quantity and quality of information provided to relatives enhances their satisfaction with healthcare services and their confidence in the process. Prior research has shown that consistent and clear communication from healthcare professionals reduces anxiety and stress levels among patients' families (15,25). Effective information-sharing also supports informed decision-making by relatives, leading to more positive ethical and clinical outcomes (28,29).

The multiple linear regression analysis revealed several key factors influencing the care satisfaction of ICU patients' relatives. The model's explanation of 36% of the variance indicates that the examined variables substantially impact satisfaction. Most notably, the level of knowledge about delirium ($\beta=0.42$, $p<0.001$) emerged as the strongest predictor, suggesting that enhancing family awareness could markedly improve satisfaction. This finding accentuates the pivotal role of the knowledge of relatives in the ICU experience (30,31). Educational attainment ($\beta=0.31$, $p=0.002$) was another significant factor, with university-educated individuals reporting higher satisfaction; which likely reflects the positive influence of education on information processing and engagement with healthcare processes.

The significant positive effects associated with the presence of a neurological disease ($\beta=0.28$, $p=0.004$) and female sex ($\beta=0.25$, $p=0.008$) suggest that satisfaction levels may be higher in contexts where family-centered care is emphasized and where comprehensive information and support are provided to relatives (30,32). The adoption of a family-centered approach in the neurology ICU where the study was conducted may have contributed to this finding by better informing and supporting relatives.

The identification of female gender as a factor increasing care satisfaction may be linked to observed tendencies among women to engage more deeply in the care process and communicate more effectively with healthcare providers. As supported by existing literature, women often assume a more active role in patient care, enabling them to navigate the ICU experience more consciously and derive greater satisfaction from the services received (21,31). The negative effect associated with a hospitalization duration of 10–13 days ($\beta=-0.22$, $p=0.018$) suggests that prolonged ICU stays may contribute to caregiver burnout and heightened stress, ultimately reducing satisfaction levels (28,32).

The analysis also indicated that younger age (20-30 years) and being the child of the patient had statistically significant, though relatively modest, positive effects on care satisfaction. The potentially greater comfort younger individuals have with digital information resources, their ability to communicate effectively with staff, and their possibly more flexible expectations could contribute to higher reported satisfaction in this group (24,30). The positive effect associated with being the patient's child may be explained by the strong emotional bonds children often have with their parents, leading to heightened emotional investment and consequently higher satisfaction when their expectations for support and care are met (28,32).

This study establishes a strong positive association between knowledge of delirium and family satisfaction among relatives of ICU patients ($r=0.60$, $p<0.001$). Regression analysis identified delirium knowledge as the foremost predictor of satisfaction ($\beta=0.42$), alongside education level, gender, and the presence of neurological disease. These results strongly advocate for the development and implementation of structured education and support programs targeted at family members.

Study Limitations

Several limitations of this study should be acknowledged when interpreting its findings. The cross-sectional design precludes the establishment of causal relationships between family members' knowledge of delirium and their care satisfaction. As the research was conducted within a single tertiary care center, the generalizability of the results to other healthcare institutions or different cultural contexts may be limited. The reliance on self-reported questionnaires introduces the potential for recall bias and social desirability bias. Furthermore, the absence of data pertaining to the actual incidence and severity of delirium among the patients may have influenced the perceptions and experiences reported by family members. Future investigations employing multicenter, longitudinal designs and incorporating objective clinical metrics are warranted to validate and extend upon these findings.

Conclusion

This cross-sectional study demonstrates that knowledge levels regarding delirium among relatives of intensive care patients are generally inadequate and that this deficit exerts a significant negative impact on family care satisfaction. The findings indicate that improving awareness and understanding of delirium represents a key modifiable factor capable of enhancing satisfaction with care.

Our results support the necessity of integrating standardized educational modules on delirium for families into routine ICU protocols and fortifying family-centered care frameworks with specific components addressing delirium management. Specifically, targeted educational programs focusing on delirium prevention, recognition, and management are anticipated to yield higher care satisfaction scores.

For future research, randomized controlled trials are recommended to evaluate the causal nature of this relationship and to assess the generalizability of these findings across diverse clinical and cultural settings.

Ethics

Ethics Committee Approval: This study received approval from the Non-Interventional Clinical Ethics Committee of İnönü University (desicion number: 2024-6/1, date: 14.01.2024). Formal institutional permission was also obtained from the administration of the Turgut Özal Medical Center.

Informed Consent: Voluntary informed consent was provided for participation.

Footnotes

Financial Disclosure: The author declared that this study received no financial support.

References

1. Pandharipande PP, Ely EW, Arora RC, Balas MC, Boustani MA, La Calle GH, et al. The intensive care delirium research agenda: A multinational, interprofessional perspective. Intensive Care Med. 2017;431329-39.
2. Li J, Fan Y, Luo R, Wang Y, Yin N, Qi W, et al. Family involvement in preventing delirium in critically ill patients. Int J Nurs Stud. 2025;81:891-905.
3. MacLullich AMJ, Beaglehole A, Hall, RJ, Meagher DJ. Delirium and long-term cognitive impairment. Int J Nurs Stud. 2019;95:1-9.
4. Rosa RG, Tonietto TF, da Silva DB, Gutierres FA, Ascoli AM, Madeira LC, et al. Perceptions about delirium in the ICU. J Clin Nurs. 2021;30:712-20.
5. Zaal IJ, Slooter AJC. Recognition of delirium in ICU patients. Intensive Crit Care Nurs. 2021;62:102967.
6. Morandi A, Pandharipande P, Trabucchi M, Rozzini R, Mistraletti G, Trompeo AC, et al. Delirium terminology in critically ill patients. Nurs Crit Care. 2017;22:203-10.
7. Jones SF, Pisani MA. ICU delirium update. Am J Crit Care. 2019;28:199-206.
8. Trogrlić Z, van der Jagt M, Bakker J, Balas MC, Ely EW, van der Voort PH, et al. ICU delirium management strategies. Crit Care Nurse. 2020;40:26-35.
9. Johnson PA, Andrews T, Stukenborg GJ, Dorman T, Rowan K. Family satisfaction with ICU care. Heart Lung. 2021;50:87-93.
10. White DB, Angus DC, Shields AM, Buddadhumaruk P, Pidro C, Paner C, et al. Improving family satisfaction in ICU. N Engl J Med. 2022;386:245-54.
11. Davidson JE, Aslakson RA, Long AC, Puntillo KA, Kross EK, Hart J, et al. Family-centered care guidelines. Crit Care Nurse. 2017;37:20-2.
12. Salluh JIF, Wang H, Schneider EB, Nagaraja N, Yenokyan G, Damluji A, et al. Outcome of delirium in ICU. BMJ. 2019;350:h2538.
13. Hwang DY, Yagoda D, Perrey HM, Tehan TM, Guanci M, Ananian L, et al. Family satisfaction with ICU care. J Nurs Care Qual. 2018;33:265-71.
14. Kynoch K, Chang A, Coyer F, McArdle A. Interventions for ICU families. Nurse Educ Today. 2019;76:15-21.
15. Duong J, Wang G, Lean G, Slobod D, Goldfarb M. Family-centered interventions in ICU. Intensive Crit Care Nurs. 2024;70:103212.
16. Fumis RR, Ranzani OT, Martins PS, Schettino G. Emotional disorders in ICU families. PLoS One. 2017;12:e0169131.
17. Abdul Halain A, Tang LY, Chong MC, Ibrahim NA, Abdullah KL. Psychological distress in ICU families. J Clin Nurs. 2022;31:497-507.
18. Ellberg C, Trieu M, Malhotra A, LaBuzetta J. Delirium and family satisfaction. Am J Nurs. 2025;125:36-43.
19. Krewulak KD, Bull MJ, Wesley Ely E, Davidson JE, Stelfox HT, Fiest KM. ICU family education on delirium. Can J Nurs Res. 2020;52:189-98.
20. Erbay Dallı O, Kelebek Girgin N. Turkish adaptation of CIDKQ. Nurs Crit Care. 2024;29:234-41.
21. Heyland DK, Tranmer JE. Measuring family satisfaction in ICU. J Nurs Meas. 2001;9:123-36.

22. Wall RJ, Engelberg RA, Downey L, Heyland DK, Curtis JR. FS-ICU survey refinement. Res Nurs Health. 2007;30:459-68.
23. Tastan S, Iyigun E, Ayhan H, Kılıckaya O, Yılmaz AA, Kurt E. Turkish version of FS-ICU. Int J Nurs Pract. 2014;20:320-6.
24. Secunda KE, Kruser JM. Patient-centered care in ICU. Crit Care Nurs Clin. 2022;34:179-88.
25. Joo Y, Jang Y, Kwon OY. Family-centered care interventions. Nurs Outlook. 2024;72:101894.
26. Fortunatti CP, Munro CL. Family satisfaction in ICU. Aust Crit Care. 2022;35:604-11.
27. Azoulay E, Pochard F, Kentish-Barnes N, Chevret S, Aboab J, Adrie C, et al. Anxiety in ICU families. Am J Respir Crit Care Med. 2020;201:156-64.
28. Scheunemann LP, Ernecoff NC, Buddadhumaruk P, Carson SS, Hough CL, Curtis JR, et al. Communication strategies for families. JAMA Nurs. 2023;2:e234567.
29. Boettger S, Nuñez DG, Meyer R, Richter A, Fernandez SF, Rudiger A, et al. Delirium review in ICU. Intensive Crit Care Nurs. 2016;35:1-8.
30. Alasad J, Ahmad M. Delirium impact on families. J Clin Nurs. 2015;24:707-15.
31. Khan BA, Perkins AJ, Campbell NL, Fox GC, Weinstein EJ, Nazir A, et al. Delirium in hospitalized patients. J Gerontol Nurs. 2020;46:22-9.
32. Sharieff S, Sajjal A, Idrees A, Rafai W. Patient and family satisfaction in the intensive care unit of a quaternary care center. Cureus. 2023;15:8-10.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.75318
Eurasian J Emerg Med. 2026;25: 53-61

Investigation of the Relationship Between Serum Subfatin Levels and Clinical Outcome in Patients with Transient Ischemic Attack in the Emergency Department

Serhat Çakır[1], Melih Yüksel[1], Mehmet Oğuzhan Ay[1], Yeşim İşler[1], Halil Kaya[1], Oğuz Ören[2], Zülfi Engindeniz[1], Ömer Furkan Demir[3], Demet Yıldız[4], Kağan Huysal[5], Ayşen Zeybek[1]

[1]University of Health Sciences Türkiye, Bursa Faculty of Medicine, Bursa Yüksek İhtisas Training and Research Hospital, Clinic of Emergency Medicine, Bursa, Türkiye
[2]Mudanya State Hospital, Clinic of Emergency Medicine, Bursa, Türkiye
[3]University of Health Sciences Türkiye, Bursa Yüksek İhtisas Training and Research Hospital, Clinic of Cardiology, Bursa, Türkiye
[4]University of Health Sciences Türkiye, Bursa Faculty of Medicine, Bursa Yüksek İhtisas Training and Research Hospital, Clinic of Neurology, Bursa, Türkiye
[5]University of Health Sciences Türkiye, Bursa Faculty of Medicine, Bursa Yüksek İhtisas Training and Research Hospital, Clinic of Clinical Biochemistry, Bursa, Türkiye

Abstract

Aim: We hypothesized that lower serum subfatin levels at admission might be associated with a higher risk of subsequent stroke in transient ischemic attack (TIA) patients. Therefore, this study aimed to examine the relationship between baseline subfatin levels and the occurrence of new cerebrovascular events at 28 and 90 days.

Materials and Methods: Patients who were admitted to the emergency department and diagnosed with TIA between 01.04.2022 and 31.03.2023 were studied prospectively.

Results: A total of 141 volunteers (71 patients and 70 controls) who met the criteria were included in the study. The median subfatin level of the patients was 1.51 [interquartile range (IQR): 25-75: 1.27-1.71] and the median subfatin level of the control group was 1.62 (IQR: 25-75: 1.13-2.32). There was a statistically significant association between coronary artery disease (CAD) and stroke development at 28 and 90 days [($p<0.05$), ($p<0.05$)]. Median subfatin levels were numerically lower in patients who experienced stroke within 28 and 90 days compared to those without subsequent stroke; however, these differences did not reach statistical significance ($p>0.05$). Therefore, no definitive conclusion can be drawn regarding the prognostic role of subfatin in this cohort.

Conclusion: CAD was significantly associated with stroke occurrence at both 28 and 90 days, underscoring its role as a major predictor of adverse outcome in TIA patients. We also found that although subfatin levels were lower in patients who had a stroke within 28 and 90 days compared to patients who did not have a stroke, there was no statistically significant difference between them.

Keywords: Emergency department, transient ischemic attack, subfatin, stroke

Introduction

Transient ischemic attack (TIA) is defined as a clinical picture of sudden onset and neurological deficits resulting from a transient decrease in blood flow to a specific area of the brain. It usually resolves spontaneously within 24 hours and is distinguished from stroke by these characteristics (1). However, the risk of stroke is significantly higher in individuals with TIA compared to the general population. The risk of stroke within the first 48 hours of TIA varies between 10-15% and this rate increases even more in the first week (2). Therefore, early evaluation, risk assessment and clinical follow-up of patients

Corresponding Author: Oğuz Ören MD, Mudanya State Hospital, Clinic of Emergency Medicine, Bursa, Türkiye
E-mail: orenoguz@gmail.com **ORCID ID:** orcid.org/0000-0002-5546-7420
Cite this article as: Çakır S, Yüksel M, Ay MO, İşler Y, Kaya H, Ören O, et al. Investigation of the relationship between serum subfatin levels and clinical outcome in patients with transient ischemic attack in the emergency department. Eurasian J Emerg Med. 2026;25: 53-61.

Received: 09.05.2025
Accepted: 06.10.2025
Epub: 23.10.2025
Published: 26.01.2026

©Copyright 2026 The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License.

with TIA is critical to prevent more serious complications such as stroke (3).

In recent years, inflammation, oxidative stress and vascular dysfunction have been recognized to play important roles in the pathophysiology of TIA and stroke. In this context, the effects of adipokines (adipose tissue-derived hormones) on the central nervous and cardiovascular systems are of increasing interest (4). Adipokines are involved in various biological processes associated with cardiovascular diseases, obesity and metabolic syndrome. One adipokine, subfatin (C1q/tumor necrosis factor-related protein 13), is involved in many metabolic processes such as energy homeostasis, inflammation and insulin sensitivity. It has also been reported that subfatin levels may be associated with atherosclerosis and cardiovascular diseases (5). However, the relationship between subfatin levels and risk factors for TIA and stroke has not been fully elucidated and research on this subject is limited.

Investigating the possible effects of subfatin on the process of brain ischemia may contribute to the prediction of stroke risk in patients with TIA. It is thought that subfatin levels may affect neuroinflammation, oxidative stress, and endothelial dysfunction, processes that play critical roles in ischemic events (6). Therefore, examining the association of biomarkers such as subfatin with stroke risk may contribute to the identification of individuals at high risk after TIA and the development of strategies that can guide the management of these patients. This study aimed to investigate the relationship between subfatin levels and the clinical outcome of patients with TIA.

Materials and Methods

Place, Time and Type of Study

This study was conducted in the emergency department of University of Health Sciences Türkiye, Bursa Yüksek İhtisas Training and Research Hospital with the approval of the clinical research ethics committee with the (desicion number: 2011-KAEK-25 2022/03-08, date: 23.1.2011), patients with a final emergency department diagnosis of TIA were prospectively analyzed.

Inclusion and Exclusion Criteria

Inclusion Criteria

Patient group;

1. Patients over 18 years of age,
2. Patients with normal radiologic imaging results,
3. Patients who gave consent to the study by themselves or their relatives were included.

In the control group, patients over 40 years of age, who had no comorbidities, and gave consent to the study, were included.

Exclusion Criteria in the Study

Patient group;

1. Patients younger than 18 years of age,
2. Patients who did not give consent to the study,
3. Patients with infarction detected on radiologic imaging,
4. Patients with bleeding detected on radiologic imaging,
5. Patients with a mass detected on radiologic imaging,
6. Pregnant patients,
7. Those with previous cerebrovascular disease were not included in the study.

In the control group, those under 40 years of age, pregnant women, and those with any comorbidities were excluded.

Work Plan

A standardized study data entry form was created for the data of the patients included in the study. Demographic information (age, gender), height, weight, body mass index, date of presentation to the emergency department (ED), vital signs (fever, respiratory rate in minutes, fingertip oxygen saturation in room air and with oxygen supplementation, systolic blood pressure), data such as diastolic blood pressure, complaints at admission, chronic diseases, medications and radiologic imaging, and the patient's outcome in the ED (discharge, admission, intensive care unit admission, excitus, treatment refusal) were recorded. In addition, patients or their relatives were contacted on days 28 and 90 and asked whether they had a new stroke attack. After the study was completed, the data in the study forms obtained were saved in electronic format for statistical analysis.

Laboratory Studies

Complete blood count, serum electrolytes (Na, Cl), renal function tests (blood urea nitrogen and creatinine), cardiac troponin, international normalized ratio levels were obtained from the patients participating in the study. To study METRNL (subfatin) level, 3 mL blood samples were collected into aprotinin tube (tube containing 500 KIU aprotinin to protect blood pretolysis).

Blood samples were allowed to clot for 30 minutes and then centrifuged at 3000 rpm for 10 minutes at room temperature. The serum obtained was portioned and stored at -80 degrees Celsius for subfatin level studies.

Human METRNL levels were determined by ELISA method in the biochemistry laboratory of S University of Health Sciences Türkiye, Bursa Yüksek İhtisas Training and Research Hospital. The ELISA kit used was Shanghai Biotechnology Co. Brand, produced in Yunnan, People's Republic of China, and the unit for Human METRNL is ng/mL. Baoshan, Romer, ChroMate 4300 ELISA reader, manufactured in Getzdersof /Austria, was used. ELISA method was performed according to the kit procedure as follows:

- Standards were prepared as recommended in the procedure. For this study, double standards and double blinds were used in each kit.
- 50 µL of the blind and standards were added to the appropriate wells. Then 40 µL of sample sera were added to the appropriate wells.
- Then 10 µL of biotin-METRNL antibody was added to the sample wells.
- 50 µL of Str-HRP-Conjugate Reagent was added to the standard and sample wells. No HRP solution was added to the blinds. The wells were then covered with a cover plate, shaken gently and incubated in an oven at 37 °C for 60 min.
- After removing from the oven, the sample was emptied. Then, 350 µL of 30 times diluted washing solution was added to each well and washed 5 times.
- After washing, 50 µL of cromogen A solution was added to each well. Then 50 µL of choromogen B solution was added to each well in dim light. The wells were covered with a cover plate, shaken gently and incubated in the dark at 37 °C for 10 min in an oven.
- The reaction was terminated by adding 50 µL of stop solution to each well. Measurement was performed at 450 nm on a ChroMate-4300 microplate reader.

Statistical Analysis

IBM SPSS Statistics for Windows, Version 21.0 (IBM Corp. Armonk, NY: USA. Released 2012) package program was used for statistical analyses. Descriptive statistics were expressed as mean ± standard deviation, median and range and/or interquartile range (IQR) for numerical variables, while categorical variables were expressed as number of cases and (%). Kolmogorov-Smirnov test was used for normality distribution of the data. Levene's test was used to determine whether the assumption of homogeneity of variances was met. The significance of the difference between the groups in terms of continuous numerical variables for which parametric test statistical assumptions were met was examined by Student's t-test. Significance of the difference in terms of continuous numerical variables for which parametric test statistical assumptions were not met was evaluated by Mann-Whitney U test. Kruskal-Wallis test was used for comparisons of three or more groups. Pearson correlation analysis was used to evaluate the relationships between parametric distributed data and Spearman's correlation analysis was used to evaluate the relationships between non-parametric distributed data. Fisher's exact test was used to analyze whether there was a relationship between categorical variables. $P<0.05$ was considered statistically significant. Results were given at 95% confidence interval.

Results

A total of 141 volunteers, including 71 patients and 70 controls, who met inclusion criteria were included in the study. The mean age of the patient group was 64.15±13.11 years and the mean age of the control group was 50.13±11.20 years. Forty-one (57.7%) of the patient group and 50 (71.4%) of the control group were male. It was found that 58 (81.7%) of the patients had comorbidities and the most common comorbidities were hypertension (n=35, 49.3%) and diabetes mellitus (n=26, 36.6%). Stroke developed within 28 days in 12 (16.9%) and within 90 days in 14 (19.7%) patients (Table 1).

The median subfatin level of the patients was 1.51 (IQR: 25-75: 1.27-1.71) and the median epicardial adipose tissue thickness was 0 mm (IQR: 25-75: 0-0.6), while the median subfatin level of the control group was 1.62 (IQR: 25-75: 1.13-2.32) and the median epicardial adipose tissue thickness was 0.3 mm (IQR: 25-75: 0.2-0.4).

While there was no statistically significant difference between the median subfatin levels of the patient group and the control group ($p>0.05$), the median epicardial adipose tissue thickness was found to be statistically significantly different ($p<0.001$) (Table 2).

In the analysis performed to determine the relationship between gender, comorbidities, medication use, electrocardiogram (ECG) findings, antiaggregant and anticoagulant use and stroke development in 28 days, a statistically significant relationship was found between coronary artery disease (CAD) and stroke development in 28 days ($p<0.05$). The stroke rate within 28 days was higher in patients with CAD (Table 3).

In the analysis performed to determine the relationship between gender, comorbidities, medication use, ECG findings, antiaggregant and anticoagulant use and stroke development in 90 days, a statistically significant relationship was found between CAD and stroke development in 90 days ($p=0.005$).

Table 1. Clinical and demographic information		
Patient group age (years)*		64.15±13.11
Control group age (years)*		50.13±11.20
Patient group gender#	Male	41 (57.7)
	Female	30 (42.3)
Control group gender#	Male	50 (71.4)
	Female	20 (28.6)
Comorbidities#		58 (81.7)
Comorbidities#	Hypertension	35 (49.3)
	Diabetes mellitus	26 (36.6)
	Coronary artery disease	18 (25.4)
	Arrhytmia	2 (2.8)
	Malignity	2 (2.8)
	Valvular diease	2 (2.8)
	Congestive heart failure	1 (1.4)
	Asthma/COPD	4 (5.6)
	Chronic kidney disease	1 (1.4)
	Others	16 (22.5)
Use of medication for comorbidity#		52 (73.2)
Antiaggregant drug use#		19 (26.8)
Anticoagulant drug use#		2 (2.8)
Electrocardiography findings #	Normal sinus rhythm	52 (73.2)
	Atrial fibrillastion	9 (12.7)
	Left bundle branch block	6 (8.5)
	Others	4 (5.6)
Emergency department outcome#	Admission	67 (94.4)
	Others	4 (5.6)
Stroke within 28 days		12 (16.9)
Stroke within 90 days		14 (19.7)
Total number of patients#		71 (100)

#n (%), *Mean ± SD, COPD: Chronic obstructive pulmonary disease, SD: Standard deviation

The rate of stroke within 90 days was found to be higher in patients with CAD (Table 4).

In the analysis performed to investigate whether there was a difference between subfatin levels and stroke status of the patients within 28 and 90 days, no statistically significant difference was found between subfatin levels and stroke status of the patients within 28 and 90 days (Table 5).

Discussion

Stroke is one of the leading causes of mortality and morbidity worldwide. TIA is a real neurological emergency due to the increased risk of stroke in patients after TIA. Many imaging methods and scoring systems are used to assess the risk of stroke after TIA. There are studies showing that these diagnostic methods are inadequate (2). In this study, we evaluated the effectiveness of serum subfatin level in determining the risk of stroke after TIA.

Hypertension is the most common comorbid disease in participants diagnosed with TIA. In our study, 81.7% of the patients had comorbid diseases and the most common comorbid disease was hypertension with a rate of 49.3%. In the study conducted by Wilson et al. (7) hypertension was found to be the most common comorbid disease with 73.5%. In a study by Kapral et al. (8) hypertension was found to be the most common comorbid disease with a rate of 66%. Our findings in this regard are consistent with other studies in the literature.

Table 2. Clinical and laboratory data of participants			
Variables	**Patient group**	**Control group**	**p value**
Age, years[&]	64.15±13.11	50.13±11.20	**<0.001**[*]
Subfatin level[ε]	1.51 (1.27-1.71)	1.62 (1.13-2.32),	>0.05[#]
Height, cm[&]	166.65±8.20	170.79±7.80	**=0.002**[*]
Weight, kg[&]	74.76±14.50	77.64±12.33	>0.05[*]
BMI[&]	26.85±4.33	26.55±4.37	>0.05[*]
Fever, C[ε]	36.2 (36.2-36.4)	36.3 (36.2-36.5)	>0.05[#]
Pulse /mi [&]	85.82±11.81	78.83±7.58	**<0.001**[*]
SBP mm/Hg[&]	145.48±22.55	127.30±6.20	**<0.001**[*]
DBP mm/Hg[&]	83.42±15.32	77.96±7.37	**=0.008**[*]
Oxygen saturation[ε]	97 (96-98)	97 (96-98)	>0.05[#]
Respiratory rate /min[ε]	14 (13-15)	12 (12-13)	**<0.001**[#]
Epicardial adipose tissue thickness/mm	0 (0-0.6)	0.3 (0.2-0.4)	**<0.001**[#]
Left ventricle end – systolic diameter/cm	3.68±0.70	2.80±0.28	**<0.001**[*]
Left ventricle end – diastolic diameter/cm	4.59±0.55	4.38±0.37	**<0.05**[*]
Left atrium diameter/cm	3.49±0.38	3.32±0.28	**=0.003**[*]
Ascending aorta diameter/cm	3.38±0.30	3.22±0.29	**=0.002**[*]
Ejection fraction[ε]	55 (50-60.0)	60 (60-60)	**<0.001**[#]
$ABCD^2$ score[ε]	4 (3-5)		
Leukocyte count[&]	8871.1±2649.9		
Neutrophil count[&]	5649.0±2217.9		
Lymphocyte count[ε]	2350 (1580-3140)		
MPV[&]	10.61±1.25		
Platelet count[&]	244940±76574		
INR[ε]	0.97 (0.89-1.08)		
Glukose/mg/dL	122 (103-143)		
Sodium/mmol/L	138.46±2.64		
Clor/mmol/L	104,62 ±3.49		
BUN/mg/dL	17 (13-23)		
Kreatinin/U/L	0.89 (0.68-1.07)		
Troponin/ng/L	4.80 (1.87-10.05)		

[ε]Median IQR (25-75), [&]Mean ± SD, [*]Student t-test, [#]Mann-Whitney U test, BMI: Body mass index, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, MPV: Mean platelet volume, BUN: Blood urea nitrogen, INR: International normalized ratio, SD: Standard deviation, IQR: Interquartile range

This may be due to the fact that hypertension disrupts the endothelial structure and leads to atherosclerosis formation.

In our study, there was a statistically significant association between CAD and stroke development at 28 and 90 days. In the study by Amarenco et al. (9) taking those without CAD as reference, the age- and sex-adjusted risk ratio of vascular events was 2.10 (0.63-6.96) for asymptomatic coronary stenosis <50%, 4.36 (1.35-14.12) for asymptomatic coronary stenosis ≥50% and 6.86 (2.15-21.31) for known CAD (9). Another study by Robinson et al. (10) also found an increased risk of stroke in patients with a history of CAD. The findings of our study are consistent with the literature.

In untreated TIAs, the risk of stroke at 3 months can reach up to 20%. Most of this risk occurs in the first 10 days, especially in the first 2 days (2,11,12). Observational data show that rapid clinical diagnosis and immediate preventive measures reduce the 3-month stroke risk by up to 80% (13). In a multicenter study in which patients with suspected TIA or mild stroke were rapidly triaged, evaluated and treated, the 3-month stroke risk was found to be approximately 5% (12). In some studies, the risk of

Table 3. Fisher's exact test of variables with stroke within 28 days						
Variables			**Stroke within 28 days**		**Total**	**Fisher's exact test**
			No	**Yes**		
Gender	Female	n (%)	25 (83.3)	5 (16.7)	30 (100)	p>0.05
	Male	n (%)	34 (82.9)	7 (17.1)	41 (100)	
Accompanying disease	No	n (%)	12 (92.3)	1 (7.7)	13 (100)	p>0.05
	Yes	n (%)	47 (81.0)	11 (19.0)	58 (100)	
Hypertension	No	n (%)	30 (83.3)	6 (16.7)	36 (100)	p>0.05
	Yes	n (%)	29 (82.9)	6 (16.1)	35 (100)	
Coronary artery disease	No	n (%)	47 (88.7)	6 (11.3)	53 (100)	p<0.05
	Yes	n (%)	12 (66.7)	6 (33.3)	18 (100)	
Diabetes mellitus	No	n (%)	37 (82.2)	8 (17.8)	45 (100)	p>0.05
	Yes	n (%)	22 (84.6)	4 (15.4)	26 (100)	
Congestive heart failure	No	n (%)	58 (82.9)	12 (17.1)	70 (100)	p>0.05
	Yes	n (%)	1 (100)	0	1 (100)	
Arrhytmia	No	n (%)	57 (82.6)	12 (17.4)	69 (100)	p>0.05
	Yes	n (%)	2 (100)	0	2 (100)	
COPD/asthma	No	n (%)	55 (82.1)	12 (17.9)	67 (100)	p>0.05
	Yes	n (%)	4 (100)	0	4 (100)	
Malignity	No	n (%)	57 (82.6)	12 (17.4)	69 (100)	p>0.05
	Yes	n (%)	2 (100)	0	2 (100)	
Chronic renal failure	No	n (%)	58 (82.9)	12 (17.1)	70 (100)	p>0.05
	Yes	n (%)	1 (100)	0	1 (100)	
Valvular heart disease	No	n (%)	57 (82.6)	12 (17.4)	69 (100)	p>0.05
	Yes	n (%)	2 (100)	0	2 (100)	
Others	No	n (%)	48 (87.3)	7 (12.7)	55 (100)	p>0.05
	Yes	n (%)	11 (68.8)	5 (31.3)	16 (100)	
Medication use	No	n (%)	17 (89.5)	2 (10.5)	19 (100)	p>0.05
	Yes	n (%)	42 (80.8)	10 (19.2)	52 (100)	
Antiagregant	No	n (%)	45 (86.5)	7 (13.5)	52 (100)	p>0.05
	Yes	n (%)	14 (73.7)	5 (16.3)	19 (100)	
Anticoagulant	No	n (%)	57 (82.6)	12 (17.4)	69 (100)	p>0.05
	Yes	n (%)	2 (100)	0	2 (100)	
Electrocardiography	Normal sinus rhythm	n (%)	43 (82.7)	9 (17.3)	52 (100)	p>0.05
	Atrial fibrillation	n (%)	9 (100)	0	9 (100)	
	Left bundle branch block	n (%)	4 (66.7)	2 (33.3)	6 (100)	
	Others	n (%)	3 (75.0)	1 (25.0)	4 (100)	
Total		n (%)	59 (83.1)	12 (16.9)	71 (100)	

COPD: Chronic obstructive pulmonary disease

an acute ischemic stroke after TIA varies between 3.5%-10% in the first 2 days and 9.2%-17% in the first three months (14-17). In our study, the stroke rate was 16.9% within 28 days and 19.7% within 90 days. The reason why these rates were partially high may have been due to post-discharge treatment non-compliance and existing comorbidities in these patients.

Subfatin is an adipokine that has been discovered in recent years and is thought to play a role in various metabolic and inflammatory processes. This protein secreted by adipose tissue plays important roles in many disease processes such as obesity, diabetes and cardiovascular diseases. It has also been suggested that subfatin may be involved in vascular inflammation, endothelial dysfunction and atherosclerotic processes. In a study

Table 4. Fisher's exact test of variables with stroke within 90 days						
Variables			**Stroke within 90 days**		**Total**	**Fisher's exact test**
			No	**Yes**		
Gender	Women	n (%)	24 (80.0)	6 (20.0)	30 (100)	p>0.05
	Men	n (%)	33 (80.5)	8 (19.5)	41 (100)	
Accompanying disease	No	n (%)	12 (92.3)	1 (7.7)	13 (100)	p>0.05
	Yes	n (%)	45 (77.6)	13 (22.4)	58 (100)	
Hypertension	No	n (%)	30 (83.3)	6 (16.7)	36 (100)	p>0.05
	Yes	n (%)	27 (77.1)	8 (22.9)	35 (100)	
Coronary artery disease	No	n (%)	47 (88.7)	6 (11.3)	53 (100)	p<0.05
	Yes	n (%)	10 (55.6)	8 (44.4)	18 (100)	
Diabetes mellitus	No	n (%)	36 (80.0)	9 (20.0)	45 (100)	p>0.05
	Yes	n (%)	21 (80.8)	5 (19.2)	26 (100)	
Congestive heart failure	No	n (%)	56 (80.0)	14 (20.0)	70 (100)	p>0.05
	Yes	n (%)	1 (100)	0	1 (100)	
Arrhytmia	No	n (%)	55 (79.7)	14 (20.3)	69 (100)	p>0.05
	Yes	n (%)	2 (100)	0	2 (100)	
COPD/asthma	No	n (%)	53 (79.1)	14 (20.9)	67 (100)	p>0.05
	Yes	n (%)	4 (100)	0	4 (100)	
Malignity	No	n (%)	55 (79.7)	14 (20.3)	69 (100)	p>0.05
	Yes	n (%)	2 (100)	0	2 (100)	
Chronic renal failure	No	n (%)	56 (80.0)	14 (20.0)	70 (100)	p>0.05
	Yes	n (%)	1 (100)	0	1 (100)	
Valvular heart disease	No	n (%)	55 (79.7)	14 (20.3)	69 (100)	
	Yes	n (%)	2 (100)	0	2 (100)	p>0.05
Others	No	n (%)	47 (85.5)	8 (14.5)	55 (100)	p>0.05
	Yes	n (%)	10 (62.5)	6 (37.5)	16 (100)	
Medication use	No	n (%)	17 (89.5)	2 (10.5)	19 (100)	p>0.05
	Yes	n (%)	40 (76.9)	12 (23.1)	52 (100)	
Antiagregant	No	n (%)	44 (84.6)	8 (15.4)	52 (100)	p>0.05
	Yes	n (%)	13 (68.4)	6 (31.6)	19 (100)	
Anticoagulant	No	n (%)	55 (79.7)	14 (20.3)	69 (100)	p>0.05
	Yes	n (%)	2 (100)	0	2 (100)	
Electrocardiography	Normal sinus rhythm	n (%)	42 (80.8)	10 (19.2)	52 (100)	p>0.05
	Atrial fibrillation	n (%)	8 (88.9)	1 (11.1)	9 (100)	
	Left bundle branch block	n (%)	4 (66.7)	2 (33.3)	6 (100)	
	Others	n (%)	3 (75.0)	1 (25.0)	4 (100)	
Total		n (%)	57 (80.3)	14 (19.7)	71 (100)	

COPD: Chronic obstructive pulmonary disease

Table 5. Analysis of the variables with subfatin levels				
Variables		**N**	**Subfatin level#**	**p value***
Stroke within 28 days	No	59	1.53 (1.23-1.73)	>0.05
	Yes	12	1.45 (1.33-1.58)	
Stroke within 90 days	No	57	1.51 (1.22-1.72)	>0.05
	Yes	14	1.47 (1.38-1.61)	

*Mann-Whitney U test, #Median interquartile range (25-75)

by Cavli et al. (18) investigating obesity/insulin resistance and subfatin levels, a significant relationship was found between low serum subfatin levels and obesity/insulin resistance. In another study by Köse et al. (19), the relationship between subfatin and preeclampsia was investigated and serum subfatin level was found to be significantly lower in preeclampsia patients compared to the control group. In the study by Demir et al. (20) investigating the relationship between ischemic stroke and serum subfatin level, no significant relationship was found. In the study by Yilmaz et al. (21) investigating subfatin levels in patients with acute coronary syndrome, subfatin levels were found to be lower in patients with non-ST elevation myocardial infarction and ST elevation myocardial infarction compared to the control group (21). In a study by Albayrak et al. (22), subfatin levels were found to be lower in patients with cerebral ischemia, intracerebral hemorrhage and subarachnoid hemorrhage compared to the control group.

In our study, although no significant difference was found between the plasma subfatin values of the patients and their stroke status on days 28 and 90, the mean serum subfatin values of the patients who had a stroke were lower than the mean serum subfatin values of the patients who did not have a stroke.

In conclusion, no significant relationship was found between serum subfatin levels and stroke in patients diagnosed with TIA in the ED. While a decrease in subfatin level poses a risk for conditions such as inflammation, endothelial damage and plaque formation, the lack of statistical difference in subfatin levels between the patient group and the control group in our study may be explained by the fact that no pathology was detected on imaging in the patient group diagnosed with TIA. The fact that serum subfatin levels were found to be low in clinical conditions such as obesity, insulin resistance and preeclampsia, which have been previously reported in the literature, suggests that this molecule may be used as a negative acute phase reactant in the future. More significant results may be obtained as the research on subfatin accumulates. Our study contributes to the literature as being the first study investigating the relationship between TIA and subfatin levels.

Study Limitations

The single-center nature of our study restrained the number of patients included. In addition, the relatively short follow-up duration of the patients, such as stroke status on days 28 to 90, may have prevented a clear demonstration of the relationship between subfatin and TIA. As control group formed with healthy individuals without known comorbidities, mean age of the control group was significantly lower than that of patient group. This is also one of the limitations of our study.

Conclusion

In our study, we found that CAD increased the risk of stroke in patients with TIA. We also observed that there was no significant difference between the subfatin levels of the patient and control groups. Finally, there was no significant difference between the subfatin levels measured at the time of admission in patients with TIA who had a stroke within 28 and 90 days and those who did not. Further research, conducted in a multi – center design, with larger patient groups and a longer follow – up period is needed on this inquiry.

Ethics

Ethics Committee Approval: This study was conducted in the emergency department of University of Health Sciences Türkiye, Bursa Yüksek İhtisas Training and Research Hospital with the approval of the clinical research ethics committee with the (desicion number: 2011-KAEK-25 2022/03-08, date: 23.1.2011).

Informed Consent: Patients who gave consent to the study by themselves or their relatives were included.

Footnotes

Author Contributions

Surgical and Medical Practices: S.Ç., M.Y., Y.İ., O.Ö., Z.E., K.H., A.Z., Concept: S.Ç., M.Y., M.O.A., Y.İ., O.Ö., Ö.F.D., K.H., A.Z., Design: S.Ç., M.Y., Y.İ., H.K., O.Ö., Z.E., Ö.F.D., D.Y., A.Z., Data Collection or Processing: S.Ç., M.Y., M.O.A., Z.E., Ö.F.D., D.Y., A.Z., Analysis or Interpretation: S.Ç., M.Y., M.O.A., Y.İ., H.K., Ö.F.D., D.Y., Literature Search: S.Ç., M.Y., O.Ö., Z.E., D.Y., K.H., Writing: S.Ç., M.Y., M.O.A., Y.İ., Z.E., K.H., A.Z.

Conflict of Interest: The authors declare that they have no conflict of interest.

Financial Disclosure: Support was received from the University of Health Sciences Türkiye, Bursa Yüksek İhtisas Training and Research Hospital Training Planing Board Scientific Research and Project Support fund.

References

1. Easton JD, Saver JL, Albers GW, Alberts MJ, Chaturvedi S, Feldmann E, et al. Definition and Evaluation of Transient Ischemic Attack: A Scientific Statement for Healthcare Professionals from the American Heart Association/American Stroke Association Stroke Council; Council on Cardiovascular Surgery and Anesthesia; Council on Cardiovascular Radiology and Intervention; Council on Cardiovascular Nursing; and the Interdisciplinary Council on Peripheral Vascular Disease. The American Academy of Neurology affirms the value of this statement as an educational tool for neurologists. Stroke. 2009;40:2276-93.
2. Johnston SC, Gress DR, Browner WS, Sidney S. Short-term prognosis after emergency department diagnosis of TIA. JAMA. 2000;284:2901-6.

3. Rothwell PM, Giles MF, Flossmann E, Lovelock CE, Redgrave JN, Warlow CP, et al. A simple score (ABCD) to identify individuals at high early risk of stroke after transient ischaemic attack. Lancet. 2005;366:29-36.
4. Rasouli N, Kern PA. Adipocytokines and the metabolic complications of obesity. J Clin Endocrinol Metab. 2008;93(11 Suppl 1):64-73.
5. Li ZY, Zheng SL, Wang P, Xu TY, Guan YF, Zhang YJ, et al. Subfatin is a novel adipokine and unlike meteorin in adipose and brain expression. CNS Neurosci Ther. 2014;20:344-54. Epub 2014 Jan 7.
6. Fenzl A, Kiefer FW. Brown adipose tissue and thermogenesis. Horm Mol Biol Clin Investig. 2014;19:25-37.
7. Wilson G, Sharma M, Eagles D, Nemnom MJ, Sivilotti MLA, Émond M, et al. Ninety-day stroke or transient ischemic attack recurrence in patients prescribed anticoagulation in the emergency department with atrial fibrillation and a new transient ischemic attack or minor stroke. J Am Heart Assoc. 2023;12:e026681.
8. Kapral MK, Hall R, Fang J, Austin PC, Silver FL, Casaubon LK, et al. Predictors of hospitalization in patients with transient ischemic attack or minor ischemic stroke. Can J Neurol Sci. 2016;43:523-8.
9. Amarenco P, Lavallée PC, Labreuche J, Ducrocq G, Juliard JM, Feldman L, et al. Coronary artery disease and risk of major vascular events after cerebral infarction. Stroke. 2013;44:1505-11.
10. Robinson K, Katzenellenbogen JM, Kleinig TJ, Kim J, Budgeon CA, Thrift AG, et al. Large burden of stroke incidence in people with cardiac disease: a linked data cohort study. Clin Epidemiol. 2023;15:203-11.
11. Amarenco P. Transient ischemic attack. N Engl J Med. 2020;382:1933-41.
12. Amarenco P, Lavallée PC, Labreuche J, Albers GW, Bornstein NM, Canhão P, et al. One-year risk of stroke after transient ischemic attack or minor stroke. N Engl J Med. 2016;374:1533-42.
13. Rothwell PM, Giles MF, Chandratheva A, Marquardt L, Geraghty O, Redgrave JN, et al. Effect of urgent treatment of transient ischaemic attack and minor stroke on early recurrent stroke (EXPRESS study): a prospective population-based sequential comparison. Lancet. 2007;370:1432-42. Erratum in: Lancet. 2008;371:386.
14. Wu CM, McLaughlin K, Lorenzetti DL, Hill MD, Manns BJ, Ghali WA. Early risk of stroke after transient ischemic attack: a systematic review and meta-analysis. Arch Intern Med. 2007;167:2417-22.
15. Ay H, Arsava EM, Johnston SC, Vangel M, Schwamm LH, Furie KL, et al. Clinical- and imaging-based prediction of stroke risk after transient ischemic attack: the CIP model. Stroke. 2009;40:181-6.
16. Chandratheva A, Geraghty OC, Luengo-Fernandez R, Rothwell PM. ABCD2 score predicts severity rather than risk of early recurrent events after transient ischemic attack. Stroke. 2010;41:851-6. Epub 2010 Mar 18.
17. Johnston SC, Rothwell PM, Nguyen-Huynh MN, Giles MF, Elkins JS, Bernstein AL, et al. Validation and refinement of scores to predict very early stroke risk after transient ischaemic attack. Lancet. 2007;369:283-92.
18. Cavli C, Önalan E, Yakar B, Dönder E, Buran İ, Önalan E. Low serum levels of meteorin-like/subfatin is related to obesity and insulin resistance. Family Practice and Palliative Care. 2022;7:137-41.
19. Köse C, Körpe B, Firatligil Fb, Aktaş Reis Y, Balkaş G, Şahin B, et al. Association between subfatin level and preeclampsia. Journal of Experimental & Clinical Medicine. 2023;40:614-8.
20. Demir HA, Seyhanli ES, Demir TG, Koyuncu I. Subfatin levels and thiol balance in patients with stroke. Neurology Asia. 2023;28:251-7.
21. Yilmaz M, Cagri Goktekin M, Ilhan N. Subfatin concentration decreases in acute coronary syndrome. Biochem Med (Zagreb). 2022;32:020704. Epub 2022 Apr 15.
22. Albayrak S, Aydin MA, Ugur K, Hanbeyoglu O, Aydin S, Erol E, et al. Subfatin, asprosin, alamandine and maresin-1 in cerebral ischemia, intracranial and subarachnoid hemorrhages. Eur Rev Med Pharmacol Sci. 2023;27:4471-80.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.67625
Eurasian J Emerg Med. 2026;25: 62-71

Quantitative Assessment of Lung and Infiltration Volumes on Chest CT for Predicting Clinical Severity and Mortality in Viral Pneumonia: Insights from COVID-19 Experience

Burak Demirci, Enes Ferhatlar, Abuzer Coşkun, Selman Gündoğan

University of Health Sciences Türkiye, İstanbul Bağcılar Training and Research Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye

Abstract

Aim: This study aimed to evaluate the relationship between lung volume, infiltration volume, and percentage of involvement with laboratory findings, hospitalization status, and mortality in patients with pneumonia.

Materials and Methods: This retrospective and observational study was conducted on 125 patients who presented to the emergency department between January and June 2021 and had pulmonary infiltration detected on chest computed tomography. The patients' lung volume, infiltration volume, and percentage of involvement were assessed along with laboratory parameters, hospitalization level, and mortality status.

Results: Among the 125 patients included, 56.8% were male and 43.2% were female, with a mean age of 49.8±14.4 years. Severity categories were as follows: mild in 85 (68.0%), moderate in 23 (18.4%), and severe in 17 (13.6%) patients. The mean total lung volume was 3034.4±1271.9 mL overall, 3398.7±1346.7 mL in the mild group, and 2192.9±466.8 mL in the severe group ($p<0.001$). The mean infiltration volume was 271.4±309.3 mL overall, 97.2±77.9 mL in the mild group, 456.7±162.9 mL in the moderate group, and 891.9±231.1 mL in the severe group ($p<0.001$). Mortality rates were 1.2% in the mild group, 30.4% in the moderate group, and 70.6% in the severe group, showing a direct association between mortality and the severity of involvement ($p<0.001$). Age, severity of involvement, C-reactive protein (CRP), and troponin levels were positively correlated with mortality.

Conclusion: In pneumonia cases, reduced lung volume, increased infiltration volume, and a higher percentage of lung involvement are predictive of mortality. Additionally, elevated CRP and troponin levels show a significant correlation with both the percentage of involvement and mortality.

Keywords: Emergency department, lung volume, pneumonia, involvement volume, mortality, COVID-19

Introduction

The chest imaging features associated with this lethal disease are crucial for achieving accurate and early diagnosis of coronavirus disease 2019 (COVID-19) (1). However, accurately predicting the impact of a disease on a patient's life, which includes both the likelihood of mortality and the severity of the illness, can be difficult. This can result in a variety of outcomes, including the absence of symptoms and severe and potentially fatal consequences. The diagnostic approach is based on a combination of factors, including a patient's history of exposure, clinical features, reverse transcription polymerase chain reaction (RT-PCR) testing, chest X-ray (CXR), and computed tomography (CT) of the thorax (2).

Microbiological testing, specifically real-time polymerase chain reaction (RT-PCR) or sequencing methodologies, is the gold standard for COVID-19 verification. Samples from the respiratory tract are used to conduct these procedures (3). CXR can be employed as the initial imaging method in COVID-19 pneumonia because it is simpler to clean the device, allows for bedside application, and contains lower radiation doses. Nevertheless,

Corresponding Author: Burak Demirci Assoc. Prof. University of Health Sciences Türkiye, İstanbul Bağcılar Training and Research Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye
E-mail: drburakdemirci@hotmail.com **ORCID ID:** orcid.org/0000-0001-6658-7260

Cite this article as: Demirci B, Ferhatlar E, Coşkun A, Gündoğan S. Quantitative assessment of lung and infiltration volumes on chest CT for predicting clinical severity and mortality in viral pneumonia: Insights from COVID-19 experience. Eurasian J Emerg Med. 2026;25: 62-71.

Received: 12.08.2025
Accepted: 12.10.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

its sensitivity is lower than that of CT because low densities are challenging to detect and may be considered normal in the early stages (4). The chest CT is employed as a diagnostic instrument for COVID-19 patients to evaluate the severity of the disease, identify diagnostic challenges, and select the most suitable treatment approach. COVID-19 pneumonia is frequently characterized by bilateral multifocal peripheral ground-glass opacities, as evidenced by a chest CT scan (5). In comparison to RT-PCR, chest CT imaging may provide a more reliable, convenient, and rapid method for the diagnosis and evaluation of COVID-19 (6).

The detailed examination of tomographic imaging, which has made a significant contribution to diagnosis and treatment, has been necessary over time. Hospitals are able to prioritize medical efforts, particularly when human resources are scarce, by classifying the severity of COVID-19 in at-risk patients. Numerous studies have developed severity score systems to evaluate the extent of COVID-19 pulmonary involvement (7-10). Although these scores enhanced the assessment of COVID-19 severity, they were not without their limitations. They require time due to their complexity. Evaluation is difficult, due to the extensive scoring in specific scores that range from 20 to 40 regions. Secondly, the right lung is larger than the left lung in overall size, not in terms of chambers or segments. Dedicated software and a trained operator are necessary even for quantitative methodologies that quantify pneumonic lesion volume (11,12).

This study aimed to illustrate the influence of this involvement percentage (IP) on clinical and mortality outcomes in light of the previous findings. This was achieved by calculating the ratio of the involvement volume to the total volume of the patient's lung in a more consistent manner, irrespective of the segment, lobe, or side.

Materials and Methods

The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki and with the consent of the Institutional Ethics Committee (protocol code: 129, date: 15.03.2021) prior to its commencement.

Study Design and Data

This retrospective and observational study was conducted in the emergency department of a tertiary training and research hospital between January 1 and June 30, 2021. During this period, a total of 2911 consecutive adult patients (≥18 years) who underwent both RT-PCR testing for SARS-CoV-2 and chest CT at their initial emergency department presentation were screened for eligibility.

According to the World Health Organization definition, only patients with a positive RT-PCR result were considered confirmed COVID-19 cases. Therefore, RT-PCR–negative patients were excluded prior to eligibility assessment. Subsequently, a series of exclusion criteria were applied to ensure data consistency and homogeneity of the study population.

Hospitalization level (general ward vs. intensive care unit) and mortality outcomes were extracted from the institutional electronic medical record system and verified with patient follow-up documentation. Quantitative CT parameters—including lung volume (LV), infiltration volume (IV), and [IP = (IV/LV) × 100]—were recorded for each patient. For interpretability, LV was categorized as 1000-2000, 2000–3000, 3000-4000, and >4000 mL, and IP was used to classify involvement severity (IS) as mild (0-15%), moderate (15-30%), and severe (>45%).

Inclusion and Exclusion Criteria

Patients aged 18 years or older with a positive RT-PCR test result for SARS-CoV-2 and evidence of parenchymal infiltration on chest CT—manifesting as ground-glass opacity and/or consolidation—were included in the study. Only those with complete demographic, laboratory, imaging, and outcome data were eligible for analysis. Patients were excluded if they had a negative RT-PCR result for SARS CoV-2 (n=1387), no parenchymal infiltration on CT (n=534), or incomplete or poor-quality imaging and/or missing laboratory data, including cases with motion artifacts or inadequate CT coverage (n=428). Additionally, individuals with known chronic comorbidities such as chronic obstructive pulmonary disease, congestive heart failure, malignancy, or chronic kidney disease (n=268), as well as those under 18 years of age (n=45) or duplicate/repeat hospital admissions (n=37), were excluded from the analysis.

After applying all exclusion criteria, 1.212 RT-PCR–positive patients remained eligible for inclusion. From this refined cohort, 125 patients were consecutively included according to the predefined inclusion criteria until the target sample size was reached (Figure 1). This approach ensured that cases were selected consecutively and without arbitrary sampling, thereby representing a clinically and temporally unbiased subset of COVID-19 pneumonia patients during the study period.

Power Analysis

Prior to analysis, a sample size justification was performed using G* Power version 3.1. The primary endpoint was in-hospital mortality, with quantitative CT metrics (IV and IP) as the main predictors. A logistic regression framework was adopted, assuming:

Two-tailed α=0.05

Statistical power (1-β)=0.80

Expected effect size [odds ratio (OR) ≈ 1.8 per SD change in IP]

Anticipated mortality rate=15-20%, consistent with prior emergency department data

Based on these assumptions, the minimum requisite sample size was calculated to be between 110 and 130 individuals, which corresponds to our final cohort of 125 patients. A post hoc power analysis utilizing the observed area under the ROC curve [area under the curve (AUC) =0.92, where IV predicted mortality] confirmed a power exceeding 0.90, hence validating the statistical sensitivity for the study's primary purpose. The study population and sample size were considered methodologically sound and aligned with established norms for retrospective observational research.

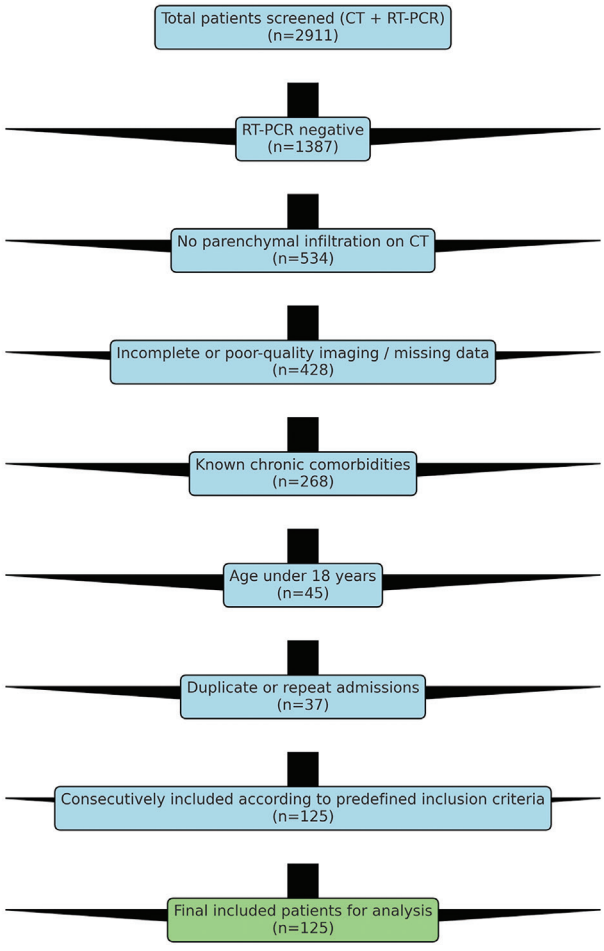


Figure 1. Patient selection flow diagram

CT: Computed tomography, RT-PCR: Reverse transcription polymerase chain reaction

Laboratory Analysis

Laboratory data were obtained from the institutional electronic medical record system, including white blood cell (WBC) count, creatinine (CRE), alanine aminotransferase, aspartate aminotransferase (AST), C-reactive protein (CRP), and troponin (TRP) values for each case. Complete blood count analyses were performed using the Sysmex DI-60 Hematology Analyzer (Sysmex Corp., Kobe, Japan), and biochemical parameters were measured with the Beckman Coulter AU-680 Automated Chemistry Analyzer (Beckman Coulter, Inc., Fullerton, CA, USA). All hematologic and biochemical results were processed and reported within approximately 45-60 minutes after sample collection.

CT Imaging and Volume Calculation

All chest CT exams were conducted using a Siemens SOMATOM Edge 128-slice multidetector CT scanner (Siemens Healthcare GmbH, Erlangen, Germany). Scans were conducted in the supine position during a single inspiratory breath-hold. The imaging settings included collimation of 128 × 0.6 mm, a rotation duration of 1.0 s, a pitch of 0.8, a tube voltage of 120 kVp, and automatic tube current modulation. Image reconstruction was executed with a 5 mm slice thickness and a 3 mm interval, utilizing a high-resolution reconstruction kernel (B70f). The CTDIvol measured 4.3 mGy, and the DLP was 175 mGy·cm.

All volumetric analyses were performed using Siemens Syngo software. Employing the VB30 workstation, utilizing the "lung analysis" and "MM reading" modules for semi-automated three-dimensional segmentation. The lung parenchyma algorithm automatically segmented each lung according to voxel density distribution. Subsequent to automatic segmentation, the radiologist meticulously adjusted the contours to exclude significant arteries, bronchi, and pleural effusions as required.

The IV was measured using a voxel-based density thresholding method, wherein all lung voxels within the attenuation range of –750 to –300 hounsfield units (HUs) were classified as exhibiting inflammatory involvement (ground-glass opacity or consolidation). The software automatically calculated for each patient:

LV: The aggregate volume of aerated lung parenchyma (in mL).

IV: The total volume of voxels designated as infiltration within the specified HU range (in mL).

IP is determined using the formula, IP = (IV/LV) x100, which indicates the fraction of lung parenchyma impacted by inflammation.

All measures were conducted by two radiologists who have 10 years of expertise in thoracic imaging and who were unaware

of the clinical results. To guarantee intra-observer reliability, each segmentation and volume extraction was conducted twice, one week apart, and the mean data were utilized for statistical analysis. Inter-observer variability was not evaluated, as the complete dataset was examined by a single reader adhering to the defined workflow methods.

After volumetric analysis, the LV, IV and [IP = (IV/LV) × 100] were automatically calculated and recorded for each patient. To facilitate clinical interpretation, LV values were categorized into four subgroups (1000-2000 mL, 2000-3000 mL, 3000-4000 mL, and >4000 mL). Based on the IP values, the IS was classified as mild (0-15%), moderate (15-30%), and severe (>45%). Hospitalization and mortality data were obtained from the institutional electronic medical records and patient follow-up forms.

Statistical Analysis

All statistical analyses were performed using IBM SPSS Statistics version 26.0 (IBM Corp., Armonk, NY, USA). Continuous variables were presented as mean ± standard deviation (SD) or median (interquartile range) according to their distribution, while categorical variables were expressed as counts and percentages. Group comparisons were made using the Student's t-test or Mann-Whitney U test for continuous variables and the χ^2 test for categorical variables. To identify factors associated with disease severity and mortality, univariate logistic regression analyses were first performed. Variables with p<0.05 in the univariate analysis were subsequently entered into a multivariate logistic regression model to control for potential confounding effects. The variables tested included age, TRP, WBC count WBC, CRP, LV, IV, and IP. Results were expressed as ORs) with 95% confidence intervals (CIs). Model performance was further assessed using ROC curve analysis, and the AUC values, were calculated to evaluate the predictive accuracy of quantitative CT parameters for mortality. A two-tailed p-value<0.05 was considered statistically significant.

Results

The mean age of the 125 patients [71 (56.8%) male, 54 (43.2%) female] included in the study was 49.78±14.40 years and there were 85 (68.0%) mild, 23 (18.4%) moderate, 17 (13.6%) severe cases in the IS groups. It was observed that IS increased with increasing age (p<0.001). The increase in CRP and TRP levels also showed a significant relationship with IS (p<0.001). While LV was determined to be 3034.37±1271.86 ml in all cases, it was 3398.71±1346.68 ml in the mild group and 2192.87±466.78 mL in the severe group (p<0.001). In the IV evaluation, the mean IV value of the cases was 271.42±309.34 ml. In the mild group, it was 97.17±77.93 mL; in the moderate group, 456.73±162.91 mL; and in the severe group, 891.99±231.09 mL (p<0.001). IP values were 3.44±3.31, 19.71±3.71, and 40.80±6.59 in the groups, respectively (p<0.001) (Table 1).

When variables were evaluated according to mortality status, the mean age in the mortality group was 62.30±12.29 years (p<0.001). High CRP and TRP levels were also significantly associated with mortality (p<0.001). In the mortality group, LV was 2441.22±1057.59 ml (p<0.017). IV and IP were 743.99±331.16 mL and 33.01±12.61 % in the mortality group and 181.41±207 (21) mL and 7.42±9.59% in the survival group. Both were significantly higher in the mortality group (p<0.001) (Table 2).

Table 1. Age, laboratory and involvement data according to the severity of involvement

		Involvement severity			
	All patients n (%): 125 (100)	**Mild n (%): 85(68.0)**	**Moderate n (%): 23 (18.4)**	**Severe n (%): 17 (13.6)**	**p value***
	Mean ± SD	**Mean ± SD**	**Mean ± SD**	**Mean ± SD**	
Age (year)	49.78±14.40	45.34±11.92	59.83±16.03	58.35±13.43	<0.001
WBC (10^3/UL)	7.26±3.75	6.47±2.65	7.64±3.20	10.66±6.51	0.027
CRE (mg/dL)	0.89±0.58	0.90±0.66	0.92±0.38	0.87±0.37	0.429
ALT (IU/L)	33.60±27.99	35.27±28.31	38.56±33.15	18.59±7.95	0.035
AST (IU/L)	43.93±32.31	44.14±26.91	56.52±51.19	25.82±10.61	0.014
CRP (mg/dL)	67.12±68.13	44.72±52.01	104.64±46.79	128.35±100.47	<0.001
TRP (ng/L)	20.33±95.86	4.83±3.68	8.38±8.26	114.03±245.45	<0.001
LV (mL)	3034.37±1271.86	3398.71±1346.68	2309.86±652.50	2192.87±466.78	<0.001
IV (mL)	271.42±309.34	97.17±77.93	456.73±162.91	891.99±231.09	<0.001
IP (%)	11.52±13.79	3.44±3.31	19.71±3.71	40.80±6.59	<0.001

*: Kruskal Wallis test, SD: Standart deviation, WBC: White blood cell, CRE: Creatinine, ALT: Alanine aminotransferase, AST: Aspartat aminotransferase, CRP: C-reactive protein, TRP: Troponin LV: Lung volume IV: Involvement volume IP: Involvement percentage

When the variables were evaluated for IS, no significance was found between gender and IS (p=0.180). In the severe group, 11 (64.7%) patients of 17 patients had LV between 2000-3000 mL, and 6 (35.3%) had LV between 1000-2000 mL. In the severe group, there was no patient with LV over 3000 mL. In the moderate group, there was no patient who had LV<4000 (p<0.001). In the mild group, there was no patient followed up in intensive care. Thirteen (56.5%) in the moderate group and 12 (70.6%) in the severe group were followed up in intensive care (p<0.001). Mortality was observed in 1 (1.2%) mild, 7 (30.4%) moderate, and 12 (70.6%) severe group cases, and increased with the severity of involvement (p<0.001) (Table 3).

Age, IS, CRP, and TRP revealed a strong positive correlation with mortality. While a strong negative correlation was determined between LV and mortality, a weak negative correlation was observed between AST and mortality (Table 4).

Table 2. Age, laboratory and involvement data according to mortality status

Mortality			
	No n (%): 105 (84.0)	**Yes n (%): 20 (16.0)**	**p value***
	Mean ± SD	**Mean ± SD**	
Age (year)	47.39±13.56	62.30±12.29	<0.001
WBC (10^3/UL)	6.65±2.60	10.45±6.49	0.053
CRE (mg/dL)	0.91±0.62	0.85±0.34	0.798
ALT (IU/L)	36.21±29.68	19.95±7.58	0.053
AST (IU/L)	47.15±34.09	27.00±9.69	0.005
CRP (mg/dL)	56.06±56.08	125.18±94.11	<0.001
TRP (ng/L)	5.08±3.59	100.43±227.74	<0.001
LV (mL)	3147.35±1282.11	2441.22±1057.59	0.017
IV (mL)	181.41±207.21	743.99±331.16	<0.001
IP (%)	7.42±9.59	33.01±12.61	<0.001

*: Mann-Whitney U tes test, SD: Standart deviation, WBC: White blood cell, CRE: Creatinine, ALT: Alanine aminotransferase, AST: Aspartat aminotransferase, CRP: C-reactive protein, TRP: Troponin LV: Lung volume IV: Involvement volume IP: Involvement percentage

Table 3. Evaluation of the involvement severity in gender, lung volume, hospitalization and mortality groups

		Involvement severity				
		Mild	**Moderate**	**Severe**	**All patients**	**p value***
		n (%)	**n (%)**	**n (%)**	**n (%)**	
Gender lung volume hospitalization mortality						
	Male	50 (58.8)	15 (65.2)	6 (35.3)	71 (56.8)	0.180
	Female	35 (41.2)	8 (34.8)	11 (64.7)	54 (43.2)	
	1000-2000	10 (11.8)	7 (30.4)	6 (35.3)	23 (18.4)	
	2000-3000	32 (37.6)	13 (56.5)	11 (64.7)	56 (44.8)	<0.001
	3000-4000	18 (21.2)	3 (13.0)	0 (0)	21 (16.8)	
	4000<	25 (29.4)	0 (0)	0 (0)	25 (20.0)	
	Inpatient service	85 (100)	10 (43.5)	5 (29.4)	100 (80.0)	<0.001
	Intensive care	0 (0)	13 (56.5)	12 (70.6)	25 (20.0)	
	No	84 (98.8)	16 (69.6)	5 (29.4)	105 (84.0)	<0.001
	Yes	1 (1.2)	7 (30.4)	12 (70.6)	20 (16.0)	
	Total	85 (100)	23 (100)	17 (100)	125 (100)	

*: Chi-square test

In the univariate logistic regression analysis, several clinical and radiological parameters—including age, TRP, CRP, IV, and IP— were found to be significantly associated with both disease severity and mortality ($p<0.05$ for each). When these variables were further evaluated in the multivariate logistic regression model, age, TRP, and IP remained independent predictors of mortality ($p<0.05$), whereas other factors lost statistical significance after adjustment for potential confounders. Among quantitative CT parameters, IP demonstrated the strongest independent association with mortality, indicating that a higher percentage of parenchymal involvement was linked to worse clinical outcomes (Table 5).

Table 4. Correlation of mortality status with variables

	Mortality	
	r	**p value***
Age	0.373	0.001
Gender	0.016	0.861
LV	-.210	0.019
IS	0.637	<0.001
WBC (10^3/UL)	0.174	0.053
CRE (mg/dL)	0.023	0.799
ALT (IU/L)	-0.174	0.052
AST (IU/L)	-0.255	0.004
CRP (mg/dL)	0.340	<0.001
TRP (ng/L)	0.362	<0.001

*Spearman's rho correlation, IS: Involvement severity, LV: Lung volume, WBC: White blood cell, CRE: Creatinine, ALT: Alanine aminotransferase, AST: Aspartat aminotransferase, CRP: C-reactive protein, TRP: Troponin

According to ROC curve analysis, the optimal cut-off values for affected volume and LV to predict mortality were identified with a threshold of 45%, yielding 97.1% sensitivity and 95.0% specificity for affected volume (AUC: 0.920; 95% CI: 0.847-0.992, $p<0.001$) and 81.9% sensitivity and 78.1% specificity for LV (AUC: 0.331; 95% CI: 0.213-0.450, $p=0.017$). (Figure 2). An IP cut off of 45% on initial chest CT predicted mortality with high accuracy (AUC: 0.92, $p<0.001$). Clinically, patients exceeding this threshold were substantially more likely to require ICU admission, suggesting that this level of pulmonary involvement may represent a critical threshold for early escalation of care.

Discussion

Pneumonia cases are frequently presented to the emergency department and are an important cause of morbidity and mortality. In pneumonia cases, the clinical course and prognosis can be affected by many factors, such as the patient's age, additional diseases, and the agent of pneumonia. Although there may be no infiltration on CT in pneumonia, we thought that the area of pulmonary involvement in cases with infiltration may also affect the clinical course. This study was motivated by the hypothesis of both LV and IV affecting the clinical outcome of cases of an atypical viral pneumonia, such as COVID-19. While evaluating these, we also had the opportunity to assess the relationship between some laboratory parameters and infiltration and clinical outcome.

Table 5. Univariate and multivariate regression analysis of involvement severity and mortality according to variables

		Univariate			Multivariate		
		OR	**95% CI**	**p value**	**OR**	**95% CI**	**p value**
Involvement severity	Age	0.445	0.020-0.870	**0.040**	0.551	0.017-0.912	**0.012**
	TRP	1.404	1.280-1.528	**<0.001**	1.392	1.261-1.613	**0.016**
	WBC	0.938	0.667-1.199	**<0.001**			
	CRP	1.113	0.954-1.271	**<0.001**			
	LV	2.141	1.834-2.448	**<0.001**			
	IV	0.876	0.806-0.946	**<0.001**			
	IP	0.878	0.829-0.927	**<0.001**	0.893	0.852-1.011	**<0.001**
Mortality	Age	1.079	1.038-1.121	**<0.001**	1.080	1.004-1.161	**0.039**
	TRP	1.216	1.103-1.340	**<0.001**	1.180	1.030-1.353	**0.017**
	WBC	1.249	1.102-1.415	**<0.001**			
	CRP	1.013	1.006-1.021	**0.001**			
	LV	0.999	0.999-1.000	**0.030**			
	IV	1.006	1.004-1.008	**<0.001**			
	IP	1.147	1.090-1.206	**<0.001**	1.331	1.046-1.695	**0.020**

OR: Odds ratio, CI: Confidence interval, WBC: White blood cell, TRP: Troponin, CRP: C-reactive protein, LV: Lung volume IV: Involvement volume IP: Involvement percentage

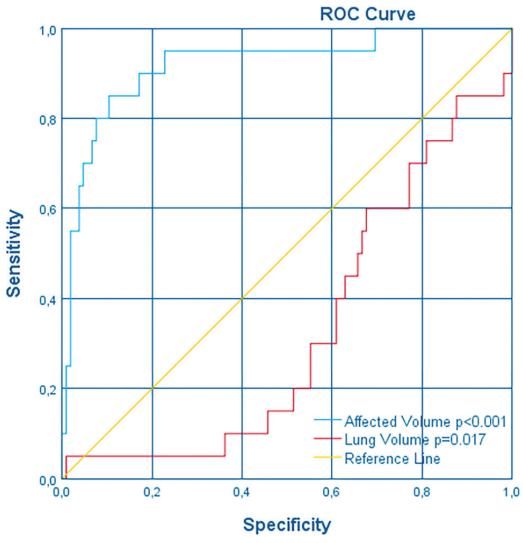


Figure 2. ROC curves showing the diagnostic performance of infiltration volume and involvement percentage for predicting mortality

Ground glass opacities are the predominant CT finding in cases of COVID-19 pneumonia. Research has documented their characteristic profiles: bilateral, peripheral, multilobar, and posterior localization (13). Ground glass opacities may be observed either independently or in conjunction with consolidation, thickening of the interlobular septum, and dilatation of blood vessels (14). Consolidations often have multifocal, segmental patchy patterns primarily observed in the lower lobe and peripheral areas, and may also include air bronchograms (15). It has been documented that they are more commonly seen in individuals of advanced age, in patients who have seen disease progression, or in the later stages of the illness (15). The experiments conducted by Pan et al. (10) revealed that consolidation is infrequent during the initial phases, but it undergoes expansion and diffusion in subsequent stages.

If we look at the main factors of age and gender, Ozdemir et al. (16) have shown in compatible studies that clinical severity increases significantly with increasing age. Regarding gender, Jin et al. (17) reported that male patients showed a worse prognosis in their study on gender factors. Similarly, in our study, we observed that both the volume involved and mortality increased significantly with increasing age. However, in our study, there was no association between gender and mortality. Unlike other studies, we did not include cases with additional diseases in our study. With this result, we can state that gender has no relationship with prognosis, even if we exclude additional diseases.

Inflammatory markers have been shown to be associated with disease severity and prognosis in COVID-19 (18,19). It is known that blood CRP and WBC elevation are associated with oxygen demand and disease severity (18,19). Many studies in the literature have reported different results regarding blood cell counts associated with the immune response. In one study, increased CRP and CRE were found to be associated with a high mortality rate. A strong relationship between increased CRP and acute lung injury in COVID-19 cases was previously shown in a study (20). Another analysis showed that CRP was higher in the group with higher severity (21). In the study by Majure et al. (22) including 6247 COVID-19 patients, TRP was a predictor of death, and significantly increased mortality rates were observed in the group of patients with high TRP levels compared to patients with normal TRP levels. Du et al. (23) studies have shown that advanced age (≥65), pre-existing concurrent cardiovascular or cerebrovascular diseases, and especially high cTnI levels of 0.05 ng/mL and above are important biomarkers for mortality in COVID-19 pneumonia patients. In our study, the relationship between some laboratory parameters and the increase in IV and mortality was evaluated. The increase in WBC, CRP, and TRP values was related to both IV and mortality, similar to findings of other studies. However, we found that CRE values were not related to these variables. This may be because we did not include chronic renal failure patients in our study, which evaluated spontaneous CRE values in cases without additional diseases.

A study by Saeed et al. (24) revealed a significant correlation between CT score and both lymphopenia and elevated levels of serum CRP, D-dimer, and ferritin. Furthermore, Zhang et al. (25) observed a positive correlation between chest CT score and CRP, erythrocyte sedimentation rate, WBC count, procalcitonin, and impaired coagulation function, and a negative correlation with lymphocyte count.

The typical findings in COVID-19 pneumonia and the changes in findings over time in the course of the disease, as well as rare and atypical findings, have been described in many studies (26). However, it should not be forgotten that similar findings can be seen in many diseases such as viral pneumonia, especially influenza, organized pneumonia, or drug toxicity. Typical CT findings and changes in findings over time have been described in the diagnosis and triage of COVID-19 pneumonia. With thoracic CT, areas with increased density, number of segments of pneumonia, and the extent to which each segment is involved can be scored using various software or visually, and the rate of lung involvement and severity of inflammation can be predicted (27). Determining severity can help predict which patients need hospitalization and mechanical ventilation. It has been reported that not only the area covered by lesions but also some findings is effective in predicting severity.

Li et al. (28) reported that a decline in the total lung capacity suggested a poor prognosis. The study conducted by Lanza et al. (29) revealed that patients with COVID-19 who required intubation had reduced total LV. Carvalho et al. (30) showed that the degradation of lung capacity directly correlated with the escalation of disease severity. In contrast to the majority of research, Ippolito et al. (31) did not demonstrate a correlation between the reduction in LV and a negative prognosis. Varying technical specifications may exist in the three-dimensional applications and devices employed in these investigations carried out at different centers. In addition, Uzun et al. (32) reported in their study that mortality increased with decreasing LV values in both sexes. In our study, consistent with previous research, both the severity and percentage of pulmonary involvement and the mortality rate were higher in patients with lower total LV. This finding can be physiologically explained by the fact that a reduction in vital capacity increases the work of breathing, leading to impaired ventilation efficiency and a higher proportional involvement of the remaining functional lung tissue.

The primary observation of COVID-19 pneumonia on thorax CT was the extensive involvement of peripheral and particularly posterior regions, as well as consolidation areas. Our findings were consistent with those documented in existing literature (33,34). Timely and precise medical evaluation of patients and proper treatment of their illnesses before they deteriorate are crucial (35). Several chest CT scoring systems were created to establish a uniform criterion for evaluating the extent of radiological lung involvement (36,37). Analysis of chest CT severity grading systems was conducted to assess interobserver agreement, revealing a high level of agreement. Nevertheless, other research is necessary to validate the findings due to the limited number of severe and critically ill patients included in both reviews. Although we created groups classifying severity as mild, moderate, and severe in the study, we think that calculating the involvement volume as a numerical percentage directly makes the results more realistic. Most studies have been conducted on the CT severity score (38), and although they show similar results to our study, our specific numerical IP calculation makes the result more valuable. As reported by Li et al. (9), the current chest CT score demonstrated a sensitivity of 80% and specificity of 82.8%, with a threshold value of 7. A semi-quantitative CT-SS devised by Pan et al. (10) was used by Francone et al. (33) to determine that a CT score of ≥18 was linked to a higher risk of mortality.

As mentioned above, previous studies have investigated the relationship between the extent of pulmonary involvement on CT and both clinical severity and mortality. In the present study, we similarly demonstrated that reduced total LV, a higher percentage of pulmonary involvement, and the overall severity of infiltration were all directly associated with increased mortality. Unlike several earlier studies that categorized disease severity using semi-quantitative CT scores, our analysis directly calculated the IP, which provided a more objective and reproducible indicator of disease burden. As the percentage of lung involvement increased, both mortality rates and ICU admissions rose proportionally. Consistent with this radiologic progression, inflammatory biomarkers such as CRP, WBC, and TRP levels also showed a significant positive correlation with the degree of parenchymal involvement.

The strongest discrimination for mortality prediction was achieved with an IP cutoff value of approximately 45% evidenced by the ROC curve analysis, which gave an AUC of 0.92. In the clinical setting, this threshold may function as an early radiological marker to identify patients who are at risk of accelerated deterioration. Patients who presented with a parenchymal involvement of ≥45% on their initial chest CT were more likely to require admission to the intensive care unit and had a significantly higher mortality rate. This observation implies that the quantitative CT assessment of lung involvement may facilitate early risk stratification and triage decisions in the emergency department. Previous imaging-based investigations have disclosed comparable results. For example, Colombi et al. (39) found that patients with more than 40-50% lung involvement on admission CT had significantly higher rates of ICU admission and mortality than those with moderate disease. Other studies have verified that quantitative CT metrics are significantly correlated with the severity of hypoxemia, inflammatory markers, and short-term outcomes in viral pneumonias, such as COVID-19 (33,40). Consequently, a clinically meaningful and reproducible cut-off of approximately 45% for IP may be used to identify patients at high risk, offering both radiologic and prognostic insight during the initial evaluation. Integrating quantitative CT assessment into emergency department workflows may help physicians identify high-risk patients earlier, guide timely escalation of care, and improve the allocation of critical care resources during surges of viral pneumonia cases.

Study Limitations

This research contains several limitations that necessitate acknowledgment. Initially, the research was conducted in a single-center, retrospective design, which may inherently restrict the generalizability of the findings and introduce selection bias. Nevertheless, a representative sample was guaranteed by the consecutive inclusion of all eligible patients in accordance with predefined criteria, thereby mitigating potential bias. Secondly, the internal validity of the results was improved by the implementation of stringent inclusion and exclusion

criteria, which in turn resulted in a more homogeneous study population, despite the fact that the sample size (n=125) was modest in comparison to the total number of screened cases. Third, only the initial chest CT scans that were obtained at the time of presentation were analyzed. As a result, the temporal evolution of lung involvement could not be assessed, and alterations in CT findings during the clinical course remain unexplored. Further longitudinal investigations that incorporate serial CT examinations are necessary to resolve this limitation. Fourth, to mitigate confounding effects, patients with substantial chronic comorbidities were excluded. Although this method restricts the external generalizability of the results to broader COVID-19 populations, it facilitated a more precise evaluation of the isolated impact of acute viral pneumonia on outcomes. Ultimately, two seasoned radiologists, each with over a decade of experience in thoracic imaging, conducted quantitative CT analyses independently. Despite the absence of formal quantification of inter-observer agreement, both readers adhered to standardized segmentation protocols and cross-validated a subset of cases to guarantee consistency. This dual-reader design significantly reduces subjective bias and improves the reliability of volumetric measurements. The present study provides a unique, quantitative, and reproducible perspective on the relationship between patient outcomes and CT-derived lung involvement, despite these limitations. The results may be a valuable reference for future multicenter prospective investigations that involve a more diverse and extensive patient population.

Conclusion

This study revealed that both quantitative increases in infiltration, and decreases in total LV on chest CT are strongly associated with adverse outcomes in viral pneumonia. A greater volume of infiltration and percentage of involvement were significantly associated with disease severity, intensive care unit hospitalization, and mortality, whereas a reduced total LV indicated advanced parenchymal damage and compromised breathing capacity. As pulmonary involvement intensified, markers of inflammation (CRP, WBC) and heart damage (TRP) increased correspondingly, highlighting a strong correlation between radiologic burden and systemic response. Among all quantitative CT characteristics, IP exhibited the most reliable predictive capability, serving as an independent indicator of worse prognosis. The findings indicate that quantitative CT assessment, encompassing both infiltration and volume analysis, might offer objective early risk classification and may inform clinical decision-making and resource allocation in emergency care.

Ethics

Ethics Committee Approval: The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki and with the consent of the Institutional Ethics Committee (protocol code: 129, date: 15.03.2021) prior to its commencement.

Informed Consent: This is a retrospective study.

Footnotes

Author Contributions

Surgical and Medical Practices: B.D., Concept: B.D., Design: B.D., E.F., Data Collection or Processing: B.D., E.F., A.C., S.G., Analysis or Interpretation: B.D., E.F., A.C., S.G., Literature Search: B.D., E.F., A.C., S.G., Writing: B.D., E.F.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Ojha V, Mani A, Pandey NN, Sharma S, Kumar S. CT in coronavirus disease 2019 (COVID-19): a systematic review of chest CT findings in 4410 adult patients. Eur Radiol. 2020;30:6129-38.
2. Lomoro P, Verde F, Zerboni F, Simonetti I, Borghi C, Fachinetti C, et al. COVID-19 pneumonia manifestations at the admission on chest ultrasound, radiographs, and CT: single-center study and comprehensive radiologic literature review. Eur J Radiol Open. 2020;7:100231.
3. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395:497-506.
4. Yoon SH, Lee KH, Kim JY, Lee YK, Ko H, Kim KH, et al. Chest radiographic and CT findings of the 2019 novel coronavirus disease (COVID-19): analysis of nine patients treated in Korea. Korean J Radiol. 2020;21:494-500.
5. Gao L, Zhang J. Pulmonary high-resolution computed tomography (HRCT) findings of patients with early-stage coronavirus disease 2019 (COVID-19) in Hangzhou, China. Med Sci Monit. 202026:e923885.
6. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology. 2020;296:E32-40.
7. Yang R, Li X, Liu H, Zhen Y, Zhang X, Xiong Q, et al. Chest CT severity score: an ımaging tool for assessing severe COVID-19. Radiol Cardiothorac Imaging. 2020;2:e200047.
8. Li K, Fang Y, Li W, Pan C, Qin P, Zhong Y, et al. CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). Eur Radiol. 2020;30:4407-16.
9. Li K, Wu J, Wu F, Guo D, Chen L, Fang Z, et al. The clinical and chest CT features associated with severe and critical COVID-19 pneumonia. Invest Radiol. 2020;55:327-31.
10. Pan F, Ye T, Sun P, Gui S, Liang B, Li L, et al. Time course of lung changes at chest CT during recovery from coronavirus disease 2019 (COVID-19). Radiology. 2020;295:715-21.

11. Chang YC, Yu CJ, Chang SC, Galvin JR, Liu HM, Hsiao CH, et al. Pulmonary sequelae in convalescent patients after severe acute respiratory syndrome: evaluation with thin-section CT. Radiology. 2005;236:1067-75.
12. Yin X, Min X, Nan Y, Feng Z, Li B, Cai W, et al. Assessment of the severity of coronavirus disease: quantitative computed tomography parameters versus semiquantitative visual score. Korean J Radiol. 2020;21:998-1006.
13. Hani C, Trieu NH, Saab I, Dangeard S, Bennani S, Chassagnon G, et al. COVID-19 pneumonia: a review of typical CT findings and differential diagnosis. Diagn Interv Imaging. 2020;101:263-8.
14. Cömert SŞ, Kıral N. COVID-19 Pnömonisinin radyolojik bulguları. South Clin Ist Euras. 2020;31:16-22.
15. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med. 2020;382:1708-20.
16. Ozdemir S, Kara BY, Gumus A, Kesimal U. The relationship between demographic and clinical characteristics of hospitalized COVID-19 patients with prognosis: cross-sectional study. Turkiye Klinikleri Arch Lung. 2022;21:51-8.
17. Jin JM, Bai P, He W, Wu F, Liu XF, Han DM, et al. Gender differences in patients with COVID-19: focus on severity and mortality. Front Public Health. 2020;8:152.
18. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020;395:1054- 62.
19. Liu A, Hammond R, Donnelly PD, Kaski JC, Coates ARM. Effective prognostic and clinical risk stratification in COVID-19 using multimodality biomarkers. J Intern Med. 2023;294:21-46.
20. Liu Y, Yang Y, Zhang C, Huang F, Wang F, Yuan J, et al. Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. Sci China Life Sci. 2020;63:364-74.
21. Shang W, Dong J, Ren Y, Tian M, Li W, Hu J, et al. The value of clinical parameters in predicting the severity of COVID-19. J Med Virol. 2020;92:2188-92.
22. Majure DT, Gruberg L, Saba SG, Kvasnovsky C, Hirsch JS, Jauhar R, et al. Usefulness of elevated troponin to predict death in patients with COVID-19 and myocardial injury. Am J Cardiol. 2021;138:100-6.
23. Du RH, Liang LR, Yang CQ, Wang W, Cao TZ, Li M, et al. Predictors of mortality for patients with COVID-19 pneumonia caused by SARS-CoV-2: a prospective cohort study. Eur Respir J. 2020;55:2000524.
24. Saeed GA, Gaba W, Shah A, Al Helali AA, Raidullah E, Al Ali AB, et al. Correlation between chest CT severity scores and the clinical parameters of adult patients with COVID-19 pneumonia. Radiol Res Pract. 2021;2021:6697677.
25. Zhang J, Meng G, Li W, Shi B, Dong H, Su Z, et al. Relationship of chest CT score with clinical characteristics of 108 patients hospitalized with COVID-19 in Wuhan, China. Respir Res. 2020;21:180.
26. Salehi S, Abedi A, Balakrishnan S, Gholamrezanezhad A. Coronavirus disease 2019 (COVID-19) imaging reporting and data system (COVID-RADS) and common lexicon: a proposal based on the imaging data of 37 studies. Eur Radiol. 2020;30:4930-42.
27. Yu M, Xu D, Lan L, Tu M, Liao R, Cai S, et al. Thin-section chest CT imaging of COVID-19 pneumonia: a comparison between patients with mild and severe disease. Radiol Cardiothorac Imaging. 2020;2:e200126.
28. Li K, Chen D, Chen S, Feng Y, Chang C, Wang Z, et al. Predictors of fatality including radiographic findings in adults with COVID-19. Respir Res. 2020;21:146.
29. Lanza E, Muglia R, Bolengo I, Santonocito OG, Lisi C, Angelotti G, et al. Quantitative chest CT analysis in COVID-19 to predict the need for oxygenation support and intubation. Eur Radiol. 2020;30:6770-8.
30. Carvalho ARS, Guimarães A, Werberich GM, de Castro SN, Pinto JSF, Schmitt WR, et al. COVID-19 chest computed tomography to stratify severity and disease extension by artificial neural network computer-aided diagnosis. Front Med (Lausanne). 2020;7:577609.
31. Ippolito D, Ragusi M, Gandola D, Maino C, Pecorelli A, Terrani S, et al. Computed tomography semi-automated lung volume quantification in SARS-CoV-2-related pneumonia. Eur Radiol. 2021;31:2726-36.
32. Uzun AY, Ucuncu Y, Hursoy N, Celiker FB, Yazici Z, et al. Mortality predictors in thorax tomography findings in COVID-19 patients. GMJ. 2024;35:149-55.
33. Francone M, Iafrate F, Masci GM, Coco S, Cilia F, Manganaro L, et al. Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis. Eur Radiol. 2020;30:6808-17.
34. Simpson S, Kay FU, Abbara S, Bhalla S, Chung JH, Chung M, et al. Radiological Society of North America expert consensus statement on reporting chest CT findings related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA - secondary publication. J Thorac Imaging. 2020;35:219-27.
35. Prokop M, van Everdingen W, van Rees Vellinga T, Quarles van Ufford H, Stöger L, Beenen L, et al. CO-RADS: a categorical CT assessment scheme for patients suspected of having COVID-19-definition and evaluation. Radiology. 2020;296:E97-104.
36. Wasilewski PG, Mruk B, Mazur S, Półtorak-Szymczak G, Sklinda K, Walecki J. COVID-19 severity scoring systems in radiological imaging - a review. Pol J Radiol. 2020;85:e361-8.
37. Guan X, Yao L, Tan Y, Shen Z, Zheng H, Zhou H, et al. Quantitative and semi-quantitative CT assessments of lung lesion burden in COVID-19 pneumonia. Sci Rep. 2021;11:5148.
38. Prakash J, Kumar N, Saran K, Yadav AK, Kumar A, Bhattacharya PK, et al. Computed tomography severity score as a predictor of disease severity and mortality in COVID-19 patients: a systematic review and meta-analysis. J Med Imaging Radiat Sci. 2023;54:364-75.
39. Colombi D, Bodini FC, Petrini M, Maffi G, Morelli N, Milanese G, et al. Well-aerated lung on admitting chest CT to predict adverse outcome in COVID-19 pneumonia. Radiology. 2020;296:E86-96.
40. Shen C, Yu N, Cai S, Zhou J, Sheng J, Liu K, et al. Quantitative computed tomography analysis for stratifying the severity of coronavirus disease 2019. J Pharm Anal. 2020;10:123-9.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.92603
Eurasian J Emerg Med. 2026;25: 72-82

Predictive Value of Electrocardiogram for Intracranial Hemorrhage After Thrombolysis in Acute Ischemic Stroke Patients: Retrospective Single-center Cohort Study

İlker Şirin[1], Nur Vahapoğlu Vural[1], Mustafa Yılmaz Alkan[1], Tuğba Sanalp Menekşe[1], Gülşah Çıkrıkçı Işık[2], Rasime Pelin Kavak[3], Erhan Saraçoğlu[4]

[1]University of Health Sciences Türkiye, Ankara Etlik City Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[2]University of Health Sciences Türkiye, Atatürk Sanatorium Training and Research Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[3]University of Health Sciences Türkiye, Ankara Etlik City Hospital, Clinic of Radiology, Ankara, Türkiye
[4]University of Health Sciences Türkiye, Ankara Etlik City Hospital, Clinic of Cardiology, Ankara, Türkiye

Abstract

Aim: Intravenous thrombolytic therapy (IVT) is an effective treatment modality for acute ischemic stroke (AIS). However, post-treatment intracranial hemorrhage (ICH), particularly symptomatic ICH (sICH), is associated with significant morbidity and mortality. This study aims to investigate the relationship between electrocardiogram (ECG)-detected left ventricular hypertrophy (LVH) and the risk of hemorrhage following IVT.

Materials and Methods: Patients who received IVT for AIS between 2022 and 2025 in a tertiary care center were retrospectively analyzed. LVH was assessed using the Gubner, Cornell, and Sokolow-Lyon ECG criteria. The primary endpoint was the development of sICH after IVT. The predictive power of ECG voltage criteria for LVH was evaluated using ROC curve analysis, and both univariable and multivariable logistic regression analyses were conducted.

Results: A total of 135 patients were included. The incidence of sICH was 11%. Increased voltage (R I + S III) according to the Gubner criterion and elevated systolic blood pressure were identified as independent risk factors for sICH ($p<0.01$). The area under the curve values for the Gubner and Cornell criteria in predicting sICH were 0.764 and 0.751, respectively. The Gubner criterion exhibited a sensitivity of 60% and a specificity of 80.8%, whereas the Cornell criterion demonstrated higher sensitivity (66.7%) but comparatively lower specificity (65.0%). The Sokolow-Lyon criterion was not statistically significant ($p=0.237$).

Conclusion: ECG-LVH may serve as a simple adjunct marker to identify patients at higher risk of sICH after IVT, but further validation is required.

Keywords: Stroke, intracranial hemorrhage, electrocardiography

Introduction

Intravenous thrombolytic therapy (IVT) with recombinant tissue plasminogen activator (rtPA) remains the cornerstone of treatment for acute ischemic stroke (AIS), particularly when administered within the therapeutic window. Despite its effectiveness in restoring cerebral perfusion, one of the most feared complications of rtPA is symptomatic intracranial hemorrhage (sICH) (1). This condition adversely affects post-treatment prognosis and constitutes a major limitation to the broader use of thrombolytic therapy. Numerous studies in the literature have investigated the predictors of post-thrombolytic hemorrhage, with cerebral microbleeds (CMB) being particularly associated with increased risk (2,3). CMB is small, round hemosiderin deposits resulting from chronic small vessel leakage, and they are detectable by magnetic resonance imaging (MRI) (4). However, due to the

Corresponding Author: İlker Şirin MD, University of Health Sciences Türkiye, Ankara Etlik City Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
E-mail: sirinilkerr@gmail.com **ORCID ID:** orcid.org/0000-0003-2694-5574

Cite this article as: Şirin İ, Vahapoğlu Vural N, Yılmaz Alkan M, Sanalp Menekşe T, Çıkrıkçı Işık G, Kavak RP, et al. Predictive value of electrocardiogram for intracranial hemorrhage after thrombolysis in acute ischemic stroke patients: retrospective single-center cohort study. Eurasian J Emerg Med. 2026;25: 72-82.

Received: 24.08.2025
Accepted: 15.10.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

need for urgent decision-making in acute stroke management, imaging is often limited to non-contrast computed tomography (CT), which restricts the evaluation of CMB.

Given the time-sensitive nature of stroke management, it is crucial that emergency physicians, neurologists, and intensive care specialists possess adequate knowledge of thrombolytic therapy and its potential complications (5). Previous studies have highlighted variability in physicians' knowledge regarding thrombolytic treatment and emphasized the importance of continuous training and protocol optimization in acute stroke care (6,7).

Although the association between CMB and increased hemorrhagic risk is well established, the inability to assess these lesions in the emergency department (ED) creates a significant diagnostic gap. Therefore, in situations requiring rapid decision-making under ED conditions, there is a need for easily accessible and applicable alternatives to advanced imaging techniques. Since chronic hypertension is the most common underlying cause of CMB, its indirect assessment through organ-based manifestations gains clinical importance in this context (8). Electrocardiography (ECG) could be a pragmatic, accessible tool, especially where MRI is not feasible. ECG-detected left ventricular hypertrophy (LVH) is the most prominent cardiac manifestation of chronic hypertension and may reflect chronic microvascular involvement. However, the association between this form of electrical remodeling and the development of ICH following IVT has not been adequately investigated in the literature.

In this retrospective cohort study, we aimed to investigate the relationship between ECG findings of LVH and the risk of both sICH and overall ICH in patients receiving IVT for AIS. If an increased risk of sICH is demonstrated among patients with LVH on ECG, this may represent a clinically valuable marker to support patient selection and ensure safer administration of IVT.

Materials and Methods

Patients and Study Design

This was a single-center retrospective cohort study conducted in a tertiary ED between 2022 and 2025, including patients diagnosed with AIS who received IVT. The study was approved by the University of Health Sciences Türkiye, Ankara Etlik City Hospital Ethics Committee (decision no: AEŞH-BADEK-2025-0230; date: 26.02.2025). It was conducted in accordance with the Declaration of Helsinki, and all patient data collected by the researchers were kept confidential.

Patients with a pre-stroke Modified Rankin scale (mRS) score greater than 2, patients with Alzheimer's disease, Parkinson's disease, or other neurodegenerative disorders, and those without available ECG data were excluded, as these conditions may cause cognitive impairment or confounding neurological deficits that could bias functional outcome assessment. In addition, patients with a history of major brain surgery, those on anticoagulant therapy, and those with inadequate brain imaging were excluded based on the predefined criteria.

Data Collection and Handling of Missing Data

Data were obtained using the hospital's electronic medical record system. Sociodemographic characteristics such as age and sex, vital signs (blood pressure, heart rate, respiratory rate, and oxygen saturation), laboratory parameters (hemoglobin, platelet count, urea, creatinine, and lactate), and patients' current medications were recorded on a standardized data collection form. mRS scores, assessed at the time of hospital admission, were also recorded (9). The etiology of ischemic stroke was classified according to the Trial of Org 10172 in acute stroke treatment criteria (10). The Alberta Stroke Program Early CT Score (ASPECTS) was calculated for the patients included in the study group (11). Additionally, three-month mortality and mRS scores at the end of the third month were obtained from the patients' electronic records.

Missing data were carefully assessed; variables with excessive missingness were excluded from the final analyses. In particular, onset-to-treatment time (OTT) was missing in the majority of patients and therefore was not included in the statistical models.

Calculation of MARS and Fazekas Scale

The Microbleed Anatomical Rating scale (MARS) was independently applied by a radiologist with 17 years of experience in neuroimaging, who was blinded to the clinical data of the patients (12). The evaluation was conducted in three stages. In the first stage, each microbleed lesion was classified as either "definite" or "possible." "Definite" lesions were defined as round, well-demarcated, hypointense foci ranging in diameter from 2 to 10 mm. "Possible" lesions were those with less distinct borders and uncertain differentiation. For lesions located in the basal ganglia, corresponding regions were carefully examined on T2-weighted and FLAIR sequences to rule out infarcts or calcifications. Vascular structures that could mimic microbleeds, imaging artifacts, air–bone interfaces, non-microbleed small hemorrhages, and mineralization foci were also meticulously evaluated.

In the second stage, microbleeds were anatomically categorized into three regions: deep, lobar, and infratentorial. Deep microbleeds included lesions located in the basal ganglia, thalamus, the internal and external capsules, corpus callosum,

and deep or periventricular white matter. Lobar microbleeds were defined as those localized to cortical or subcortical areas. Infratentorial microbleeds encompassed those identified in the brainstem and cerebellum. In the final stage, the total MARS score was calculated by summing the number of microbleeds observed in the aforementioned regions.

Leukoaraiosis (white matter lesions) was evaluated using the Fazekas scale on FLAIR MRI sequences. The Fazekas score was assigned separately for periventricular and deep white matter hyperintensities, each rated from 0 to 3. Periventricular signal abnormalities were assessed based on the presence of caps, rims, or extensive lesions, while deep white matter lesions were scored according to the number, size, and degree of confluence. The sum of both subscores constituted the total Fazekas score.

Evaluation of Electrocardiographic and Echocardiographic Left Ventricular Hypertrophy

For the assessment of LVH on ECG, three commonly used criteria were adopted: the Cornell voltage criterion (13), the Gubner criterion (14), and the Sokolow-Lyon criterion (15), each representing unipolar limb, bipolar limb, and chest leads. All ECGs were independently evaluated by three emergency medicine specialists blinded to the hemorrhagic outcomes. For each ECG, the average of the voltage measurements was calculated, and it was subsequently reviewed and confirmed by the cardiologist who was also blinded to clinical outcomes. The values approved by the cardiologist were used for statistical analysis. Although blinding was ensured, inter-rater reliability among the emergency specialists was not measured, which represents a limitation of the ECG interpretation process.

The definitions for the three voltage criteria were as follows:

Cornell voltage criterion: Male: RaVL + SV3 >28 mm (=2.8 mV), Female: RaVL + SV3 > 20 mm (=2.0 mV)

Gubner voltage criterion: RI + SIII ≥ 25 mm (=2.5 mV)

Sokolow-Lyon criterion: SV3 + RV5 or V6 > 35 mm (=3.5 mV)

In our hospital, echocardiographic (ECHO) evaluation is routinely performed within one week of hospitalization in patients diagnosed with AIS. ECHO reports and archived imaging were reviewed by a cardiologist experienced in echocardiography. LVH was recorded in the data form if the left ventricular wall thickness exceeded 11 mm in men or 10 mm in women.

Intracranial Hemorrhage Outcome

Any ICH detected by CT within 24 hours after rtPA administration was defined as post-thrombolytic ICH. In clinical practice, post-thrombolytic ICH includes both sICH and asymptomatic ICH. Symptomatic hemorrhages were classified according to the Heidelberg Bleeding Classification, which involves a structured seven-step process for categorizing hemorrhagic events (16). In this study, the Heidelberg classification was performed by more than one reviewer; however, inter-rater reliability was not assessed, which represents a limitation of the outcome definition. However, due to the inherent limitations of retrospective data review, hemorrhages were defined in this study as events associated with any of the following: ≥4-point worsening in the total National Institutes of Health Stroke scale (NIHSS) score compared to the most recent pre-deterioration assessment (not baseline); ≥2-point worsening in any individual NIHSS category (to detect new hemorrhages causing new symptoms); or the need for intubation, hemicraniectomy, external ventricular drainage, or other major medical/surgical interventions.

The anatomical classification of ICH was conducted according to the predefined criteria summarized in Table 1.

Sample Size

Based on previous studies that reported a 30% incidence of ICH in patients who received thrombolytic therapy for AIS (17), and assuming a clinically meaningful difference in the prevalence of LVH between those with and without ICH (18), it was estimated that a minimum of 24 patients with ICH would be sufficient to achieve 85% statistical power with a two-sided alpha of 0.05. The post-hoc power was computed in G* Power (Exact test; one-sample proportion, two-tailed, α=0.05) using the expected ICH incidence under H0 of 30% and the observed incidence of 17.8% with n=135. The achieved power was approximately 66%.

Table 1. Anatomical classification of intracranial hemorrhage according to the Heidelberg bleeding classification

Class	Type	Description
1		Hemorrhagic transformation of infarcted brain tissue
	HI[1]	Scattered small petechiae, no mass effect
	HI[2]	Confluent petechiae, no mass effect
	PH[1]	Hematoma within infarcted tissue, occupying <30%, no substantive mass effect
2	PH[2]	Hematoma occupying ≥30% of infarcted tissue, with significant mass effect
3		Intracerebral hemorrhage outside infarcted tissue or extracerebral bleeding
	3[a]	Parenchymal hematoma remote from infarcted tissue
	3[b]	Intraventricular hemorrhage
	3[c]	Subarachnoid hemorrhage
	3[d]	Subdural hemorrhage

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, Version 25.0 (IBM Corp., Armonk, NY, USA). Continuous variables were summarized as medians and interquartile ranges (IQRs), while categorical variables were expressed as frequencies and percentages. Between-group comparisons were performed using the Mann-Whitney U test for continuous variables and the Chi-square or Fisher's exact test for categorical variables, as appropriate.

To evaluate the predictive performance of ECG LVH criteria for sICH, ROC curve, analysis was conducted, and area under the curve (AUC), optimal cut off values, sensitivity, and specificity were reported. Univariable and multivariable logistic regression analyses were used to identify independent predictors of sICH. Variables with a p value <0.20 in univariable analysis were included in the multivariable model. A two-sided p value of <0.05 was considered statistically significant.

Results

A total of 135 patients diagnosed with AIS and treated with IVT were included in the final analysis. Figure 1 presents the patient selection process and flowchart. The mean age of the cohort was 67.2±14.8 years, and 60.7% (n=82) were male. The median systolic and diastolic blood pressures at presentation were 150.5 mmHg and 89 mmHg (IQR: 80-95), respectively. Table 2 summarizes the baseline demographic and clinical characteristics of the study population.

When stratified by the presence of sICH, no significant differences were observed in age or sex. However, systolic and diastolic blood pressures were significantly higher in the sICH group (p=0.002 and p=0.004, respectively). While ASPECTS and MARS scores did not differ significantly between groups, the 3-month mRS score was lower in the sICH group (p=0.005), likely reflecting the significantly higher mortality observed in this group (40% vs. 8%, p<0.001). There was no statistically significant difference in the rate of bridging therapy between patients with and without ICH, and between those with and without sICH. These findings are presented in Table 3. Detailed data are provided in Supplementary Table 1.

For predicting ICH, the Cornell voltage criterion demonstrated the highest [AUC: 0.641, 95% confidence interval (CI): 0.510-0.773, p=0.030], with 83.3% sensitivity and 27% specificity. The Gubner criterion also reached statistical significance (AUC: 0.628, 95% CI: 0.491–0.766, p=0.049), whereas the Sokolow-Lyon criterion did not show a significant discriminative ability (p=0.398) (Figure 2).

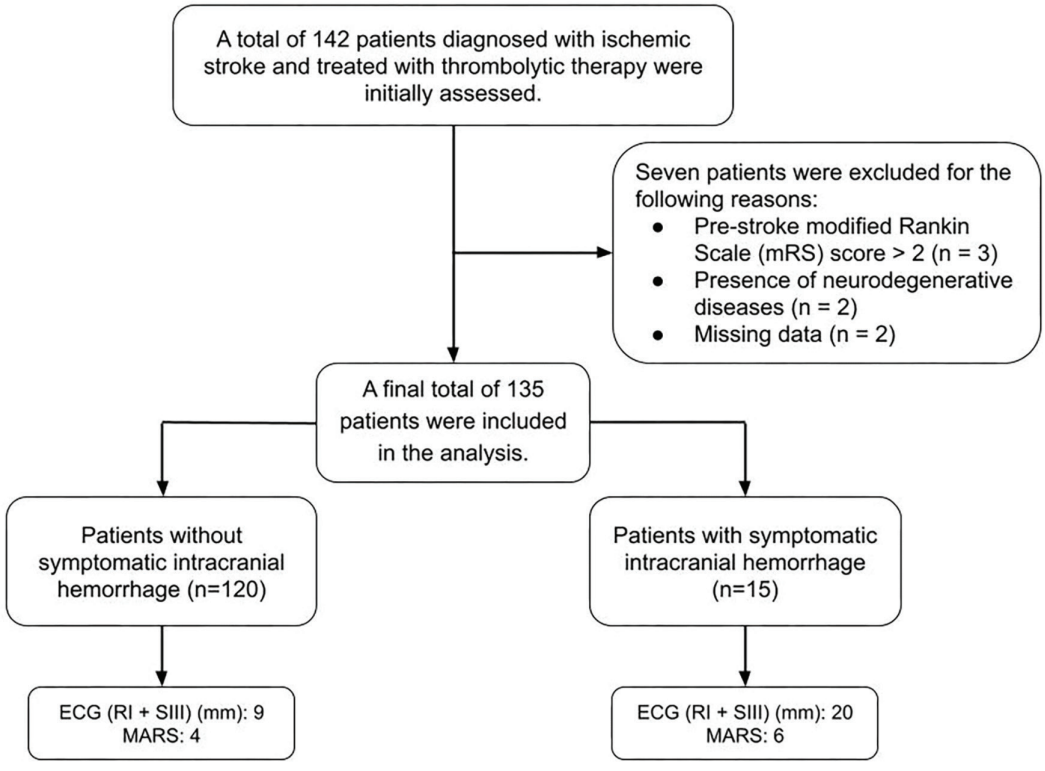


Figure 1. Flowchart of patient selection and group stratification
ECG: Electrocardiogram, MARS: Microbleed Anatomical Rating scale

Table 2. General characteristics of the study population	
Variables (n=135)	
Sex	
Female	53 (39.3%)
Male	82 (60.7%)
Age	67.19±14.78
Vital parameters	
Systolic blood pressure (mmHg)	150.5±21.96
Diastolic blood pressure (mmHg)	89 (IQR: 80-95)
Pulse (beat/minute)	76 (IQR: 69-87)
Respiratory rate (per minute)	20 (IQR: 18-20)
Oxygen saturation (%)	95 (IQR: 92-97)
Pharmacological therapies administered	
Antiplatelet therapy	59 (43.7%)
Anticoagulant therapy	10 (7.4%)
Statin therapy	15 (11.1%)
Antihypertensive therapy	84 (62.2%)
Laboratory parameters	
Hemoglobin (g/dL)	13.9±2.09
Platelet count (.103/L)	242.8±75.8
Urea level (mg/dL)	36.9 (IQR: 28-48.5)
Creatinine level (mg/ dL)	0.9 (IQR: 0.78-1.09)
Lactate level (mmol/L)	1.81 (IQR: 1.38-2.28)
Trial of Org 10172 for acute stroke treatment classification	
Large-artery atherosclerosis	85 (63%)
Cardioembolic stroke	23 (17%)
Small-vessel disease/penetrating artery disease	12 (8.9%)
Stroke of other determined cause	8 (5.9%)
Stroke of undetermined cause (cryptogenic stroke)	7 (5.2%)
Alberta stroke program early CT score	9 (IQR: 7-10)
Microbleed Anatomical Rating scale	4 (IQR: 2-6)
Fazekas scale score (n=87)	
Periventricular white matter	2 (IQR: 1-2)
0 =no lesion	11 (12.6%)
1 =“caps” or pencil-thin lining	32 (36.8%)
2 =smooth “halo”	41 (47.1%)
3 =irregular periventricular signal extending into the deep white matter	3 (3.4%)
Deep white matter	1 (IQR: 1-2)
0 =no lesion	16 (18.4%)
1 =punctate lesions	37 (42.5%)
2 =beginning confluence	31 (35.6%)
3 =extensive confluent areas	3 (3.4%)
Modified Rankin scale at administration	3 (IQR: 3-4)
Post- thrombolytic bleeding	
Yes	24 (17.8%)
Heidelberg bleeding classification (n=24)	1
Class 1: hemorrhagic transformation of infarcted brain tissue	2
1^a: HI1: scattered small petechiae, no mass effect	6
1^b: HI2: confluent petechiae, no mass effect	
1^c: PH1: hematoma within infarcted tissue, occupying <30%, no substantive mass effect	
Class 2: intracerebral hemorrhage within and beyond infarcted brain tissue	4
PH^2: hematoma occupying ≥30% of the infarcted tissue, with obvious mass effect (PH^2)	
Class 3: intracerebral hemorrhage outside the infarcted brain tissue or intracranial-extracerebral hemorrhage	
3^a: parenchymal hematoma remote from infarcted brain tissue	1
3^b: intraventricular hemorrhage	2
3^c: subarachnoid hemorrhage	7
3^d: subdural hemorrhage	1

Table 2. Continued	
Variables (n=135)	
Electrocardiographic criteria of left ventricular hypertrophy The Gubner criteria (voltage) Hypertrophic (R I + S III >25 mm) **The Sokolow-Lyon criteria (voltage)** Hypertrophic (S V3 + R V5-6 >35 mm) The Cornell Voltage Criteria (voltage) Hypertrophic (R aVL + S V3 >35 mm for men, >25 mm for women) Hypertrophic findings at echocardiography (*n=115)	 9 (IQR: 6-15) mm 11 (8.1%) 16.5 (IQR: 12-22) mm 5 (3.7 %) 14 (IQR: 9-18) mm 6 (4.4%) 38 (33%)
Length of hospital stay (days)	8 (IQR: 5-13)
Modified Rankin scale at 3rd month	3 (IQR: 1-3)
Bridging therapy (mechanical thrombectomy)	32 (24 %)
Symptomatic intracranial hemorrhage	15 (11 %)
Survey Alive Died	 119 (88.1%) 16 (11.9%)
*Number of patients that performed echocardiography. All data are presented as counts and percentages for categorical variables, and as either mean ± standard deviation or median and interquartile range (IQR: 25th-75th percentiles), depending on the distribution of the data, and adherence to normality assumptions CT: Computed tomography, IQR: Interquartile range	

Table 3. Comparison of clinical, radiological, and electrocardiographic characteristics according to the presence of intracranial hemorrhage and symptomatic intracranial hemorrhage

Variables **Outcome 1: intracranial hemorrhage**	**Yes (n=24)**	**No (n=111)**	**p value**
Sex Female Male	 12 (50%) 12 (50%)	 41 (36.9%) 70 (63.1%)	 0.235
Age	72.08±11.86	66.13±15.17	0.073
Vital parameters Systolic blood pressure (mmHg) Diastolic blood pressure (mmHg)	 164.5 (IQR: 142.5-177.5) 92.5 (IQR: 87-100)	 145 (IQR: 135159) 88 (IQR: 80-93)	 **0.004** 0.066
Microbleed Anatomical Rating scale	5 (IQR: 3-6.75)	4 (IQR: 2-6)	0.205
Survey Died	 10 (42%)	 6 (5%)	 **<0.001**
Electrocardiographic criteria of left ventricular hypertrophy The Gubner criteria (voltage) Hypertrophic (R I + S III >25 mm) The Sokolow-Lyon criteria (voltage) Hypertrophic (S V3 + R V5-6 >35 mm) The Cornell Voltage Criteria (voltage) Hypertrophic (R aVL + S V3 >35 mm for men, >25 mm for women)	 9 (IQR: 6-13) mm 6 (25%) 16 (IQR: 13-23) mm 2 (8.3%) 15 (IQR: 13-22) mm 3 (12.5%)	 6 (IQR: 12-26) mm 5 (4.5%) 17 (IQR: 12-22) mm 3 (2.7%) 13 (IQR: 9-17) mm 3 (2.7%)	 0.083 **<0.001** 0.380 0.185 0.054 **0.035**
Hypertrophic findings at echocardiography (*n=115) Hypertrophic	n=16 9 (56.3%)	n=99 29 (29.3%)	**0.033**
Variables Outcome 2: symptomatic intracranial hemorrhage	**Yes (n=15)**	**No (n=120)**	**p value**
Sex Female Male	 7 (47%) 8 (53%)	 46 (38%) 74 (62%)	 0.582
Age	71.1±11.4	66.7±15.1	0.300
Vital parameters Systolic blood pressure (mmHg) Diastolic blood pressure (mmHg)	 166 (IQR: 155-180) 100 (IQR: 89-109)	 147 (IQR: 135-160) 88 (IQR: 80-94)	 **0.002** **0.004**

Table 3. Continued			
Variables **Outcome 1: intracranial hemorrhage**	**Yes (n=24)**	**No (n=111)**	**p value**
Survey Died	 6 (40 %)	 10 (8%)	 **<0.001**
Electrocardiographic criteria of left ventricular hypertrophy The Gubner criteria (voltage) Hypertrophic (R I + S III >25 mm) The Sokolow -Lyon criteria (voltage) Hypertrophic (S V3 + R V5-6 >35 mm) The Cornell Voltage Criteria (voltage) Hypertrophic (R aVL + S V3 >35 mm for men, >25 mm for women)	 20 (IQR: 12-28) mm 5 (33%) 18 (IQR: 11-21) mm 1 (7%) 20 (IQR: 14-26) mm 3 (20%)	 9 (IQR: 5-13) mm 6 (5%) 16 (IQR: 12-24) mm 4 (3%) 13 (IQR: 9-17) mm 3 (3%)	 **0.001** **<0.001** 0.236 0.521 **0.002** **0.002**
Hypertrophic findings at echocardiography (*n=115) Hypertrophic	n=10 6 (40%)	n=105 32 (27%)	 0.281
*: Patients who underwent echocardiography (number of patients provided), CT: Computed tomography, IQR: Interquartile range			

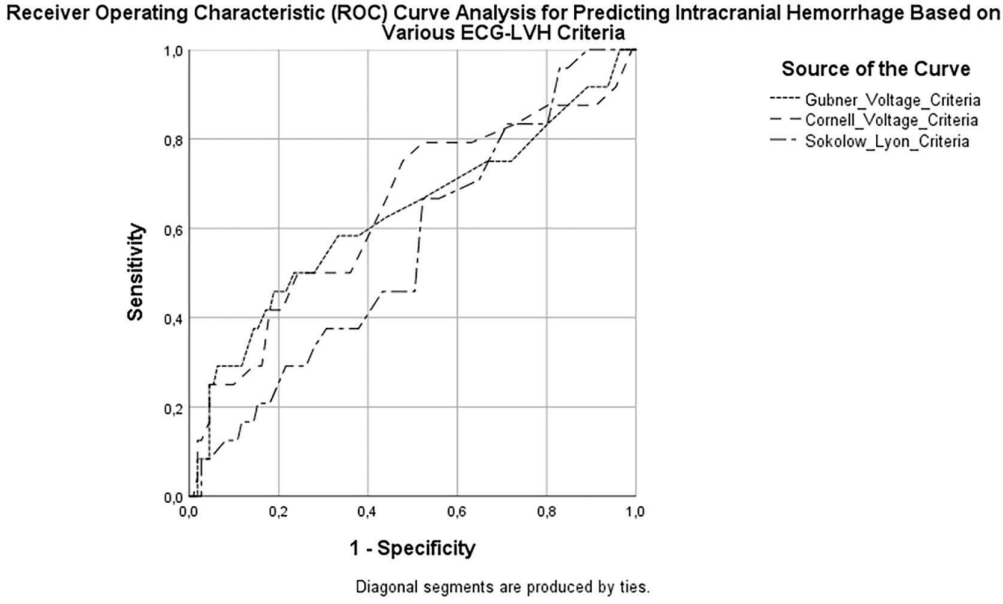


Figure 2. ROC curve analysis of ECG-LVH criteria for predicting any intracranial hemorrhage
ECG-LVH: Electrocardiogram- left ventricular hypertrophy

In predicting sICH, both the Gubner and Cornell Voltage Criteria demonstrated statistically significant performance (AUC: 0.764 and 0.751, respectively; both p<0.01) (Figure 3). These AUC values indicate moderate, rather than high, discriminative power. The Gubner criterion achieved 60% sensitivity and 80.8% specificity, while the Cornell criterion demonstrated slightly lower specificity (65.0%) but higher sensitivity (66.7%). The Sokolow-Lyon criterion again yielded nonsignificant results (AUC: 0.594, p=0.237) (Table 4).

Univariable and multivariable logistic regression analyses were conducted to identify independent predictors of sICH. In the univariable analysis, Gubner voltage [odds ratio (OR): 1.120, 95% CI: 1.045-1.201, p=0.001] and systolic blood pressure (OR: 1.039, 95% CI: 1.012-1.066, p=0.004) were significantly associated with sICH, while age was not (p=0.166).

In the multivariable model, which included Gubner voltage and systolic blood pressure, both remained independent predictors of sICH. Gubner voltage (OR: 1.126, 95% CI: 1.049-1.210, p=0.001) and systolic blood pressure (OR: 1.038, 95% CI: 1.011-1.066, p=0.006) were both independently associated with an increased risk of sICH (Table 5).

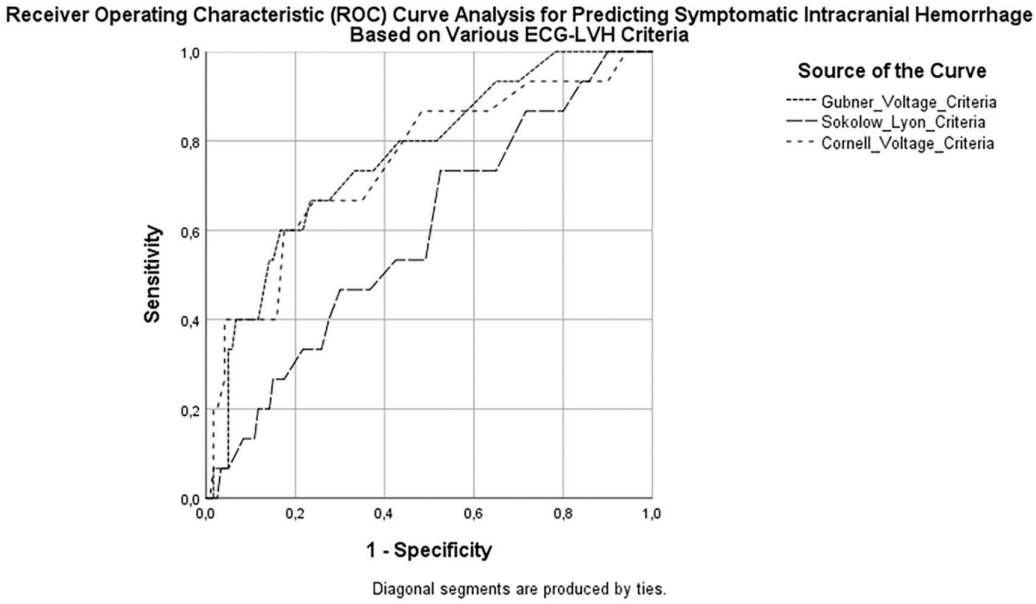


Figure 3. ROC curve analysis of ECG-LVH criteria for predicting symptomatic intracranial hemorrhage
ECG-LVH: Electrocardiogram- left ventricular hypertrophy

Tablo 4. Predictive performance of electrocardiographic left ventricular hypertrophy criteria for intracranial hemorrhage and symptomatic intracranial hemorrhage

Outcome 1: Diagnostic performance of ECG criteria for predicting intracranial hemorrhage								
LVH criteria	AUC	95% CI	p value	Best cut off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
The Gubner criteria	0.628	0.491 -0.766	**0.049**	10.5	66.7	47.7	73.3	62.5
Cornell voltage criteria	0.641	0.510 -0.773	**0.030**	14.5	83.3	27	66.7	55.9
Sokolo-Lyon criteria	0.555	0.434- 0.676	0.398	13.5	83.3	19.8	70.8	35.1
Outcome 2: Diagnostic performance of ECG criteria for predicting symptomatic intracranial hemorrhage								
The Gubner criteria	0.764	0.640- 0.888	**0.001**	15.5	60.0	80.8	32.0	93.8
Cornell voltage criteria	0.751	0.611- 0.891	**0.002**	15.5	66.7	65.0	28.0	94.5
Sokolow-Lyon criteria	0.594	0.448- 0.740	0.237	16.5	53.3	50.8	25.0	90.0

ECG: Electrocardiography, AUC: Area under the curve, CI: Confidence interval, PPV: Positive predictive value, NPV: Negative predictive value

Table 5. Univariable and multivariable logistic regression analysis of factors associated with symptomatic intracranial hemorrhage

Variable	**Univariable OR (95% CI)**	**p value**	**Multivariable OR (95% CI)**	**p value**
Gubner voltage	1.120 (1.045-1.201)	**0.001**	1.126 (1.049-1.210)	**0.001**
Systolic blood pressure	1.039 (1.012-1.066)	**0.004**	1.038 (1.011-1.066)	**0.006**
Age	1.034 (0.986-1.085)	0.166		

OR: Odds ratio, CI: Confidence interval

Discussion

In this study evaluating ECG left ventricular voltage criteria as predictors of ICH and sICH following thrombolytic therapy, in patients with AIS, the Gubner and Cornell Voltage Criteria demonstrated moderate discriminatory power (AUC values around 0.75). Moreover, the RI + SIII (Gubner voltage criterion) was identified as an independent predictor of symptomatic hemorrhage. These findings suggest that ECG-detected LVH showed moderate predictive value and may complement, rather than replace, established predictors.

Both hemorrhagic transformation (HT) and remote ICH following IVT are known to independently increase the risk of poor outcomes (19). Therefore, predicting the occurrence of any type of ICH after IVT is of clinical importance. While some risk factors are shared, distinct predictors have been identified for HT, parenchymal hematoma (PH), and remote ICH. Male sex, a history of prior stroke, OTT, and infarct size have been found to be independent risk factors for HT (20,21). On the other hand, congestive heart failure, advanced age, elevated systolic blood pressure, and atrial fibrillation are known risk factors for PH (22). Pathophysiological mechanisms underlying some of these factors have also been described. For instance, prior research has demonstrated that blood–brain barrier (BBB) permeability increases with age (23). Hypertension may impair vascular elasticity, leading to hyperperfusion injury. Markers of small vessel disease such as cerebral amyloid angiopathy, white matter hyperintensities, and CMB may predict remote ICH through endothelial dysfunction and disruption of the BBB (24,25).

In this context, MARS has been proposed as a potential tool to evaluate the functional consequences of CMBs in AIS patients treated with IVT (26). Since CMBs are often regarded as a consequence of chronic hypertension, their presence may confer predictive value for hemorrhagic complications following IVT. This could explain why mean ECG voltages were found to be significantly higher in patients who experienced post-thrombolytic hemorrhage—whether symptomatic or not—compared to those who did not.

In the current study, no significant difference in LVH was observed between sICH and non-sICH patients based on ECHO findings. However, both the proportion of patients meeting LVH criteria and the mean voltages were significantly higher in the sICH group when assessed using the Gubner and Cornell criteria. While ECHO-LVH reflects anatomical changes in left ventricular wall thickness, ECG-LVH encompasses both anatomical and electrical remodeling (27). Both ECHO-LVH and ECG-LVH have been shown to possess independent prognostic value (28). In our study, the discrepancy between ECG and ECHO-based findings should be cautiously interpreted and requires further validation, rather than being solely attributed to electrical versus anatomical remodeling.

Assessing LVH using ECG voltage criteria may help improve acute care quality in emergency settings by enabling early prediction of sICH after IVT. Although current guidelines clearly define indications for IVT in AIS patients (29), predicting post-thrombolytic hemorrhage remains a challenge. An increased risk of hemorrhage in patients with CMB has been well documented (30). However, detecting CMBs in the ED MRI, and calculating MARS scores requires radiological expertise, both of which increase time and cost. In contrast, evaluating LVH on ECG is simple, widely accessible, and does not require additional testing. Nevertheless, ECG findings should not be used in isolation for treatment decisions but may be integrated into multiparametric risk models that also include established predictors such as NIHSS, blood pressure, and imaging findings.

In the present study, the Gubner and Cornell Voltage Criteria demonstrated sufficient predictive power for sICH, whereas the Sokolow-Lyon criterion did not perform adequately. Previous research has also shown the superiority of the Cornell and Gubner criteria over the Sokolow-Lyon criterion in predicting sudden cardiac death and detecting anatomical LVH (31,32). Notably, an RI + SIII voltage sum exceeding 15.5 mm was associated with approximately 80% specificity for sICH, suggesting that even patients not formally meeting LVH criteria (e.g., RI + SIII >25 mm) based on Gubner may still carry elevated hemorrhagic risk.

The association between cardiac and cerebral injury appears to stem from the systemic effects of arterial hypertension. Both organs are recognized target of hypertension, as previously demonstrated (33). Consistent with prior studies establishing the prognostic value of LVH in cerebrovascular events (34), our findings indicate that RI + SIII on ECG serves as an independent risk factor for sICH following IVT in AIS patients.

Study Limitations

This study has several limitations. First, its retrospective design, may have introduced potential bias in data collection and, by nature, is more susceptible to confounding factors. Second, the study was conducted at a single center. However, the center where the study was performed is the largest hospital in the country and serves as a major referral center for AIS patients.

Third, although the Heidelberg Bleeding Classification was used to define hemorrhagic outcomes, comparison with findings in the existing literature may be limited, as some previous studies have utilized different classification systems, such as the European Cooperative Acute Stroke Study (ECASS) II or ECASS

III. Nevertheless, the Heidelberg classification offers a more comprehensive framework than ECASS III, and it has been shown to have good interobserver agreement (35,36).

Fourth, although ECG evaluations were independently performed by three emergency medicine specialists, inter-rater reliability was not assessed, which may limit the robustness of ECG interpretation. Fifth, the relatively small number of sICH cases (n=15) reduces statistical power and limits the generalizability of the findings.

Sixth, although the impact of OTT on post-IVT ICH is well established in AIS, it could not be analyzed in this study due to a lack of data for the majority of patients. Nonetheless, in our hospital, current guideline-based indications for IVT are strictly followed, suggesting that OTT may have had minimal influence on treatment-related complications in our cohort.

Finally, MARS scoring could not be performed in all included patients, which may have influenced the findings related to CMB. However, it is important to note that MRI prior to IVT is not mandatory in the acute stroke setting. Even so, we acknowledge this as a methodological limitation.

Conclusion

ECG-detected LVH, particularly by the Gubner and Cornell criteria, was associated with sICH after IVT. These findings are hypothesis-generating and suggest that ECG may serve as an adjunct risk marker, but larger, prospective multicenter studies are needed for validation.

Ethics

Ethics Committee Approval: The study was approved by the University of Health Sciences Türkiye, Ankara Etlik City Hospital Ethics Committee (decision no: AEŞH-BADEK-2025-0230; date: 26.02.2025). It was conducted in accordance with the Declaration of Helsinki, and all patient data collected by the researchers were kept confidential.

Informed Consent: This was a single-center retrospective study.

Generative AI Usage Statement

Generative AI and AI-assisted technologies were NOT used in the preparation of this work.

Footnotes

Surgical and Medical Practices: İ.Ş., Concept: İ.Ş., R.P.K., E.S., Design: İ.Ş., T.S.M., G.Ç.I., N.V.V., M.Y.A., Data Collection or Processing: İ.Ş., R.P.K., E.S., Analysis or Interpretation: İ.Ş., G.Ç.I., Literature Search: İ.Ş., G.Ç.I., Writing: İ.Ş., R.P.K., E.S., T.S.M., G.Ç.I., R.P.K., E.S.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Lee GR, Ko SH, Choi HS, Hong HP, Lee JS, Jeong KY. Prevalence, risk factors, and clinical outcomes of remote intracerebral hemorrhage after intravenous thrombolysis in acute ischemic stroke: a systematic review and meta-analysis. Clin Exp Emerg Med. 2024;11:171-80.
2. Oraby MI, Gomaa R, Abdel-Aal AA, Hussein M. Cerebral microbleeds in acute ischemic stroke after intravenous thrombolysis and their impact on short term outcome of stroke. Int J Neurosci. 2024:1-9.
3. Tipirneni S, Stanwell P, Weissert R, Bhaskar SMM. Prevalence and impact of cerebral microbleeds on clinical and safety outcomes in acute ischaemic stroke patients receiving reperfusion therapy: a systematic review and meta-analysis. Biomedicines. 2023;11:2865.
4. Kimura K, Aoki J, Shibazaki K, Saji N, Uemura J, Sakamoto Y. New appearance of extraischemic microbleeds on T2*-weighted magnetic resonance imaging 24 hours after tissue-type plasminogen activator administration. Stroke. 2013;44:2776-81.
5. Shams Vahdati S, Ala A, Sadeghi Hokmabadi E, Afsar E, Shiri F, et al. Assessing the emergency medicine specialists' knowledge regarding thrombolytic therapy in ischemic stroke patients. Educ Res Med Sci. 2025;13:e153456
6. Ala A, Vahdati SS, Tahmasebi P, Gilani N, Abbasi A, Aletaha R. The effect of ASLS instruction on the knowledge and skills of medical students in Tabriz University of Medical Sciences, Iran, Single group pre-and post-test design. Res Dev Med Educ. 2025;14:33294.
7. Mehdizadehfar E, Sadeghi Hokmabadi E, Taheraghdam A, Sadeghpour Y, Khabbaz A, et al. Optimizing acute ischemic stroke care: evaluating emergency stroke code activation and thrombolytic therapy in east azerbaijan province. Arch Neurosci. 2024;11:e147803.
8. Greenberg SM, Vernooij MW, Cordonnier C, Viswanathan A, Al-Shahi Salman R, Warach S, et al. Cerebral microbleeds: a guide to detection and interpretation. Lancet Neurol. 2009;8:165-74.
9. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. Stroke. 1988;19:604-7.
10. Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke. 1993;24:35-41.
11. Barber PA, Demchuk AM, Zhang J, Buchan AM. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. ASPECTS Study Group. Alberta Stroke Programme Early CT Score. Lancet. 2000;355:1670-4.
12. Gregoire SM, Chaudhary UJ, Brown MM, Yousry TA, Kallis C, Jäger HR, et al. The microbleed anatomical rating scale (MARS): reliability of a tool to map brain microbleeds. Neurology. 2009;73:1759-66.
13. Casale PN, Devereux RB, Alonso DR, Campo E, Kligfield P. Improved sex-specific criteria of left ventricular hypertrophy for clinical and computer interpretation of electrocardiograms: validation with autopsy findings. Circulation. 1987;75:565-72.
14. Gubner R, UNGERLEIDER HE. Electrocardiographic criteria of left ventricular hypertrophy: factors determining the evolution of the electrocardiographic patterns in hypertrophy and bundle branch block. Arch Intern Med . 1943;72:196-209.

15. Sokolow M, Lyon TP. The ventricular complex in left ventricular hypertrophy as obtained by unipolar precordial and limb leads. Am Heart J. 1949;37:161-86.
16. von Kummer R, Broderick JP, Campbell BC, Demchuk A, Goyal M, Hill MD, et al. The heidelberg bleeding classification: classification of bleeding events after ischemic stroke and reperfusion therapy. Stroke. 2015;46:2981-6.
17. Pande SD, Win MM, Khine AA, Zaw EM, Manoharraj N, Lolong L, et al. Haemorrhagic transformation following ischaemic stroke: a retrospective study. Sci Rep. 2020;10:5319.
18. Landolfi A, Selvetella G, Cugino D, Grillea G, Maffei A, Notte A, et al. Hemorrhagic transformation of acute ischemic stroke is limited in hypertensive patients with cardiac hypertrophy. Int J Cardiol. 2016;219:362-6.
19. Hao Y, Zhou H, Pan C, Xie G, Hu J, Zhang B, et al. Prediction factors and clinical significance of different types of hemorrhagic transformation after intravenous thrombolysis. Eur J Med Res. 2023;28:509.
20. Jensen M, Schlemm E, Cheng B, Lettow I, Quandt F, Boutitie F, et al. Clinical characteristics and outcome of patients with hemorrhagic transformation after intravenous thrombolysis in the WAKE-UP trial. Front Neurol. 2020;11:957.
21. Wen L, Zhang S, Wan K, Zhang H, Zhang X. Risk factors of haemorrhagic transformation for acute ischaemic stroke in Chinese patients receiving intravenous thrombolysis: a meta-analysis. Medicine (Baltimore). 2020;99:e18995.
22. Honig A, Percy J, Sepehry AA, Gomez AG, Field TS, Benavente OR. Hemorrhagic transformation in acute ischemic stroke: a quantitative systematic review. J Clin Med. 2022;11:1162.
23. Kaur J, Tuor UI, Zhao Z, Barber PA. Quantitative MRI reveals the elderly ischemic brain is susceptible to increased early blood-brain barrier permeability following tissue plasminogen activator related to claudin 5 and occludin disassembly. J Cereb Blood Flow Metab. 2011;31:1874-85.
24. Zhang CE, Wong SM, van de Haar HJ, Staals J, Jansen JF, Jeukens CR, et al. Blood-brain barrier leakage is more widespread in patients with cerebral small vessel disease. Neurology. 2017;88:426-32.
25. Chen Y, Yan S, Xu M, Zhong G, Liebeskind DS, Lou M. More extensive white matter hyperintensity is linked with higher risk of remote intracerebral hemorrhage after intravenous thrombolysis. Eur J Neurol. 2018;25:380-e15.
26. Li G, Wang C, Wang S, Wang L, Hao Y, Xiong Y, et al. Clinical significance of the microbleed anatomical rating scale score in ischemic stroke patients treated with intravenous thrombolysis. Postgrad Med J. 2024;100:421-6.
27. Bacharova L, Szathmary V, Kovalcik M, Mateasik A. Effect of changes in left ventricular anatomy and conduction velocity on the QRS voltage and morphology in left ventricular hypertrophy: a model study. J Electrocardiol. 2010;43:200-8.
28. Porthan K, Kenttä T, Niiranen TJ, Nieminen MS, Oikarinen L, Viitasalo M, et al. ECG left ventricular hypertrophy as a risk predictor of sudden cardiac death. Int J Cardiol. 2019;276:125-9.
29. Berge E, Whiteley W, Audebert H, De Marchis GM, Fonseca AC, Padiglioni C, et al. European Stroke Organisation (ESO) guidelines on intravenous thrombolysis for acute ischaemic stroke. Eur Stroke J. 2021;6:I-LXII.
30. Zhan Z, Fu F, Zhang W, Cheng Z. Prevalence, Risk factors, and clinical outcomes of new cerebral microbleeds after intravenous thrombolysis in acute ischemic stroke: a systematic review and meta-analysis. Clin Neuroradiol. 2024;34:209-18.
31. You Z, He T, Ding Y, Yang L, Jiang X, Huang L. Predictive value of electrocardiographic left ventricular hypertrophy in the general population: a meta-analysis. J Electrocardiol. 2020;62:14-9.
32. Courand PY, Lantelme P, Gosse P. Electrocardiographic detection of left ventricular hypertrophy: time to forget the sokolow-lyon index? Arch Cardiovasc Dis. 2015;108:277-80.
33. Bikkina M, Levy D, Evans JC, Larson MG, Benjamin EJ, Wolf PA, et al. Left ventricular mass and risk of stroke in an elderly cohort. The Framingham Heart Study. JAMA. 1994;272:33-6.
34. Selvetella G, Notte A, Maffei A, Calistri V, Scamardella V, Frati G, et al. Left ventricular hypertrophy is associated with asymptomatic cerebral damage in hypertensive patients. Stroke. 2003;34:1766-70.
35. Neuberger U, Möhlenbruch MA, Herweh C, Ulfert C, Bendszus M, Pfaff J. Classification of Bleeding Events: Comparison of ECASS III (European Cooperative Acute Stroke Study) and the New Heidelberg Bleeding Classification. Stroke. 2017;48:1983-5.
36. van der Ende NAM, Luijten SPR, Kluijtmans L, Postma AA, Cornelissen SA, van Hattem AMG, et al. Interobserver agreement on intracranial hemorrhage on magnetic resonance imaging in patients with ischemic stroke. Stroke. 2023;54:1587-92.

Supplementary Table 1. Comparison of baseline characteristics, pharmacological therapies, imaging findings, and clinical outcomes between patients who developed intracranial hemorrhage or symptomatic intracranial hemorrhage and those who did not

Variables **Outcome 1: Intracranial hemorrhage**	**Yes (n=24)**	**No (n=111)**	**p value**
Pharmacological therapies administered Antiplatelet therapy Anticoagulant therapy Antihypertensive therapy	 11 (45.8%) 4 (16.7%) 19 (79.2%)	 48 (43.2%) 6 (5.4%) 65 (58.6%)	 0.817 0.056 0.059
Alberta stroke program early CT score	8 (IQR: 7.75-10)	8 (IQR: 7-10)	0.178
Modified rankin scale at administration	4 (IQR: 3- 4)	3 (IQR: 3-4)	**0.027**
Length of hospital stay (days)	12.5 (IQR: 6.5-23)	7 (IQR: 5-12)	**0.012**
Modified rankin scale at 3rd month	4 (IQR: 3-6)	2.5 (IQR: 1-3)	**<0.001**
Bridging therapy (mechanical thrombectomy)	5 (21 %)	26 (23%)	0.784
Variables Outcome 2: symptomatic intracranial hemorrhage	**Yes (n=15)**	**No (n=120)**	**p value**
Alberta stroke program early CT score	8.5 (IQR: 7-10)	8 (IQR: 6.5-9.5)	0.558
Microbleed anatomical rating scale	6 (IQR:1.5-7.5)	4 (IQR: 2-6)	0.255
Modified rankin scale at administration	4 (IQR: 3.5-4.5)	3 (IQR: 2.5-3.5)	0.305
Length of hospital stay (days)	17.5 (IQR: 4.5-30.5)	8 (IQR: 4-12)	0.075
Modified rankin scale at 3rd month	2.5 (IQR: 1-4)	3 (IQR: 2-4)	**0.005**
Bridging therapy (mechanical thrombectomy)	4 (27%)	11 (9%)	0,747

CT: Computed tomography, IQR: Interquartile range

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.80457
Eurasian J Emerg Med. 2026;25: 83-9

Case-Based Radiology Education In Interns: Comparing Learning And Retention Across Modalities

İsmail Ataş[1], Mümin Murat Yazıcı[1], Doğa Hakkani[2], Okan Kekeç[2], Ahmet Buğra Şıvgın[2], Hayrunnisa Azman[2], Enes Güler[3], Özlem Bilir[1]

[1]Recep Tayyip Erdoğan University Training and Research Hospital, Clinic of Emergency Medicine, Rize, Türkiye
[2]Recep Tayyip Erdoğan University Faculty of Medicine, Rize, Türkiye
[3]Ministry of Health Kırklareli Provincial Health Directorate, Clinic of Emergency Medicine, Kırklareli, Türkiye

Abstract

Aim: This study aimed to evaluate intern physicians' ability to interpret basic radiological imaging methods and to investigate the effect of a case-based training program on knowledge acquisition and retention.

Materials and Methods: In this prospective study, participants underwent a pre-test before training, a post-test after training, and a retention test four weeks later. Assessments were performed using direct radiography, ultrasound (US), computed tomography (CT), and magnetic resonance imaging (MRI) modalities. The data were statistically analyzed, and performance changes between modalities were examined.

Results: This study included 50 intern physicians enrolled in medical school. Of the participants, 40% (n=20) were male and 60% (n=30) were female. The median age was 24 years (interquartile range: 24-24). A statistically significant increase was found in all imaging modalities in the post-test results compared to the pre-test (direct radiography, US, CT: $p<0.001$; MRI: $p=0.019$). In the retention test, a statistically significant increase was observed in all modalities compared to the pre-test ($p=0.001$). However, when the post-test was compared with the retention test, no statistically significant difference was observed in direct radiography and US ($p=0.381$; $p=0.059$), while a statistically significant decrease was observed in CT and MRI ($p=0.006$; $p=0.001$).

Conclusion: Case-based training significantly improves interns' ability to interpret basic radiologic imaging modalities. While permanent learning was achieved primarily in direct radiography and basic US applications, loss of knowledge was observed in more complex modalities such as CT and MRI. Structuring training programs taking into account these differences may contribute to the development of clinical decision-making skills.

Keywords: Intern physician, radiology education, case-based learning, ultrasound, retention

Introduction

Radiologic imaging modalities are an indispensable part of the diagnosis and treatment process in the emergency department. Direct radiography, computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound (US) especially facilitate rapid decision-making in emergency patient management. However, many studies have shown that intern physicians feel inadequate in the interpretation and application of these imaging modalities. In a study by Glenn-Cox et al. (1), a large proportion of intern physicians reported that they did not receive adequate radiology training in the medical school curriculum and therefore experienced deficiencies in clinical practice. Similarly, Harthoorn et al. (2) emphasized that radiology education should be integrated more effectively into the curriculum, and restructuring teaching methods and content could increase students' acquisition of clinical skills. In addition, other studies in the field of radiology have demonstrated the positive relationship between students' attitudes, skills, and clinical competencies, indicating the need for more effective planning of educational processes (3).

In recent years, online or case-based learning methods have come to the forefront alongside traditional methods in clinical education (4). Case-based learning enables students to acquire

Corresponding Author: İsmail Ataş MD, Recep Tayyip Erdoğan University Faculty of Medicine, Rize, Türkiye
E-mail: drismailatas@gmail.com **ORCID ID:** orcid.org/0000-0001-6723-8564
Cite this article as: Ataş İ, Yazıcı MM, Hakkani D, Kekeç O, Şıvgın AB, Azman H, et al. Case-based radiology education in interns: comparing learning and retention across modalities. Eurasian J Emerg Med. 2026;25: 83-9.

Received: 21.08.2025
Accepted: 17.10.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

knowledge through real clinical scenarios and to develop problem-solving and decision-making skills. Fromke emphasized that case-based learning is an effective and efficient teaching model, especially for new generation students (Gen Z and Millennial generation) (5). Sugi et al. (6) showed that the use of interactive case materials in radiology education enriches the learning process by increasing student participation. In addition, the new case-based and competency-oriented teaching format developed by Masthoff et al. (7) increased student motivation and supported skill acquisition. this study aimed to evaluate the effectiveness of a case-based theoretical and practical training program designed to improve the radiological interpretation and application skills of intern physicians. It also aimed to investigate its impact on short-term learning and long-term knowledge retention.

Materials and Methods

Study Design and Setting

This prospective study was conducted in a medical school within the scope of the Scientific and Technological Research Council of Türkiye (TÜBİTAK) 2209-A University Students Research Projects Support Program. Within the scope of the study, a case-based theoretical and practical training program on radiological imaging frequently encountered in the emergency department was implemented. The results of the pre-test, post-test, and retention test conducted after the training constituted the research data. This study was approved by the institutional review board and ethics committee of Recep Tayyip Erdoğan University in Türkiye (decision no: 2025/357, date: 31.07. 2025).

Participant Selection and Data Collection

A total of 62 intern physicians studying in a medical school, who were in the internship period, did not have a shift or outpatient clinic/service shift at the hospital on the dates of the study, and were able to participate in all sessions uninterruptedly during the training were invited to the study. Four of the intern physicians missed their theoretical training, two missed their practical training, two missed the pre-test, one missed the post-test, and three missed the retention test. As a result, out of the intern physicians were excluded, and 50 who completed all training and evaluation processes were included in the study.

Prior to the training, a questionnaire including demographic information, subjective perceptions of skills and attitudes towards radiology, and a 10-question case-based multiple-choice pre-test was administered to the participants. After the completion of the training, the 10-question post-test and questionnaire were reapplied which were prepared at equivalent content and difficulty level. To evaluate the retention of learning, a retention test consisting of 28 questions was conducted 1 month after the training. All data were recorded using a standard data coding template and converted into a format suitable for analysis.

Training Program and Content

The training program was planned for two days. On the first day, case-based theoretical training were held. In this context, lectures titled "The Importance of Direct Radiography", "Commonly Missed Direct Radiography Examples", "Computed Tomography with All Sections", "Common Pathological Computed Tomography Cases", "US, the Shining Star of Recent Years", "Common Ultrasound Case Examples" and "Magnetic Resonance imaging Reading Principles and Case Examples" were presented. In the theoretical sessions, each topic was explained through case examples of common pathologies encountered in the emergency department.

On the second day, hands-on practical training were held to improve the participants' skills in the use of US. Four stations were created in this context: "Trauma US — the Focused Assessment with Sonography in Trauma (FAST) and the extended FAST (eFAST)", "hepatobiliary US", "Doppler US", and "US-Guided Interventional Procedures (vascular access, venipuncture, central venous catheter placement)". At each station, the training sessions were supported by visual materials, video recordings, and live model applications. The training program was conducted in an integrated structure, aiming to provide both theoretical knowledge and practical skills.

All theoretical and practical training sessions were conducted by emergency medicine specialists who had completed at least five years of professional experience. For the US practice sessions, two instructors were selected, who had both attended and successfully completed a basic US course approved by professional emergency medicine associations. Both had at least five years of experience in point-of-care ultrasound (PoCUS), with an average of 250 US examinations per year.

Data Collection Tools and Measurements

The data collection process in the study consisted of three main components: a demographic and subjective assessment questionnaire, case-based knowledge tests (pre-test and post-test), and a retention test.

Questionnaires: In the questionnaires administered before and after the training, participants' gender, age, grade level, the specialty they planned to choose in the future, their subjective evaluation of the type, duration, and adequacy of their previous radiology education, their self-initiated learning activities, and their perception of their skills in interpreting imaging methods were questioned. In the questionnaires administered after the training, additional questions were included regarding the scientific adequacy of the program, its appropriateness in terms

of duration, the level of experience gained, and the imaging technique in which improvement was achieved. Likert-type scales were also used to assess the effect of the training program on the participants' ability to interpret or apply different imaging modalities. Accordingly, the participants were asked to respond on a five-point scale before and after the training program for direct radiography, CT, and MRI, by selecting "definitely can interpret, can interpret, undecided, cannot interpret, definitely cannot interpret" and for US by selecting definitely can apply, can apply, undecided, cannot apply, definitely cannot apply. Thus, an attempt was made to measure whether there was a change in individual perceptions resulting from the training.

Pre-test and post-test: Both tests consisted of 10 questions and were prepared at a similar level of difficulty. The questions were organized in a case-based multiple-choice format based on four direct radiographs, 3 CT, 2 US, and 1 MRI image. The pre-test was administered just before the start of the training, and the post-test was administered after the completion of the training. In tests, the number of correct answers was evaluated based on both total score and modality (direct radiography, CT, US, MRI).

Retention test: It was administered 1 month after the completion of the training and consisted of 28 questions. The questions were prepared in a case-based multiple-choice format including eight direct radiographs, nine CT, seven US, and four MR images. With this test, the long-term retention of the information obtained from the training was evaluated.

All test questions were developed collaboratively by three emergency medicine specialists with at least five years of clinical experience and active involvement in undergraduate radiology and US education. The content validity of the questions was ensured through consensus among the educators, focusing on common emergency radiology scenarios. To guarantee equivalence of difficulty and scope between the pre-test and post-test, the questions were constructed using identical learning objectives, similar case complexity, and parallel distributions of imaging modalities. The retention test questions were independently reviewed and validated by two external emergency medicine educators, to confirm that they assessed the same level of knowledge and comprehension without reproducing identical cases.

Statistical Analysis

All analyses were performed in Jamovi version 1.6 statistical software (The Jamovi Project, 2021, Computer Software, version 1.6, Sydney, Australia). Continuous variables were expressed as mean ± standard deviation or median interquartile range (IQR). Categorical variables were expressed as number (percentage, %). In modality-based learning outcome and retention assessments, the percentage of correct answers for the relevant modality was calculated. Pre-tests, and post-tests retention tests were compared using paired Student's t-test when parametric assumptions were met, and Wilcoxon signed-rank test, when parametric assumptions were not met. In all statistical analyses, p values <0.05 were considered significant.

Results

A total of 50 intern physicians were included in the study. Of the participants, 40% (n=20) were male, and 60% (n=30) were female. The median age was 24 years (IQR: 24-24), and the age distribution of the participants was relatively homogeneous. Urology (12%, n=6), emergency medicine (10%, n=5), cardiology (8%, n=4), and psychiatry (8%, n=4) were the most frequently considered specialties in career choices. Of the participants, 46% (n=23) reported previous didactic radiology training, 30% (n=15) had case-based training experience, and 92% (n=46) had attempted self-learning. The median values for didactic training time and case-based training time were 0 hours (IQR: 0-10) and 0 hours (IQR: 0-2), respectively. Basic demographic information is shown in Table 1.

Figure 1 shows the pre- and post-training self-assessment results of the intern physicians. While 50% of the participants stated that they could interpret direct radiographs before the training, this rate increased to 98% after the training. In US: while 74% of the participants answered "I cannot" or "I am undecided" before the training, 64% stated that they could perform the application after the training. In CT, while only 28% thought that they could interpret before the training, 80% stated that they could interpret after the training. In magnetic resonance interpretation, while 50% of the participants were undecided before the training, 52% stated that they could interpret, after the training. In general evaluation, most of the participants reported that the training made a significant contribution, especially in the fields of direct radiography and US (Figure 1).

The pre-test and post-test performances of the intern physicians participating in the study are presented in Table 2. According to the post-test results, a significant increase was found in all imaging modalities: direct radiography ($p<0.001$), US ($p<0.001$), CT ($p<0.001$), and MRI ($p = 0.019$). Similarly, in Table 3, where the results of the pre-test and retention test were compared, a significant improvement was observed in all modalities ($p=0.001$ for all). In Table 4, where post-test and retention test performances were evaluated, no significant difference was found between the direct radiography and US results ($p=0.381$, $p=0.059$, respectively). In contrast, a significant difference was found in CT and MRI ($p=0.006$, $p=0.001$, respectively). These data revealed that case-based training provided more permanent learning, especially in direct radiography and US applications.

Table 1. The volunteers' demographic data and baseline characteristics	
Characteristics	**All volunteers (n=50)**
Gender	
Male, n (%)	20 (40.0)
Female, n (%)	30 (60.0)
Age (years), median (IQR)	24 (IQR: 24-24)
Specialties considered in career preference	
Emergency medicine, n (%)	5 (10.0)
Family medicine, n (%)	2 (4.0)
Internal medicine, n (%)	2 (4.0)
Infectious diseases, n (%)	3 (6.0)
General surgery, n (%)	1 (2.0)
Eye diseases, n (%)	3 (6.0)
Thoracic surgery, n (%)	2 (4.0)
Public health, n (%)	1 (2.0)
Obstetrics and gynecology, n (%)	1 (2.0)
Cardiology, n (%)	4 (8.0)
Neurosurgery, n (%)	1 (2.0)
Orthopedics, n (%)	2 (4.0)
Pediatrics, n (%)	1 (2.0)
Psychiatry, n (%)	4 (8.0)
Medical genetics, n (%)	1 (2.0)
Anesthesia, n (%)	1 (2.0)
Dermatology, n (%)	3 (6.0)
Pulmonologist, n (%)	1 (2.0)
Plastic surgery, n (%)	1 (2.0)
Urology, n (%)	6 (12.0)
Child psychiatry, n (%)	1 (2.0)
Unknown, n (%)	4 (8.0)
Didactic education, n (%)	23 (46.0)
Didactic education (time), median (IQR)	0 (IQR: 0-10)
Case-based education, n (%)	15 (30.0)
Case-based education (time), median (IQR)	0 (IQR: 0-2)
Self-directed learning, n (%)	46 (92.0)
IQR: Interquartile Range (25p, 75p)	

Discussion

This study showed that after case-based theoretical training enriched with emergency cases and structured US practice, short-term cognitive acquisition increased significantly in all modalities for intern physicians. One month later, the data showed that the increase in direct radiography and US performance was maintained, while there was a partial decline in CT and MRI. The findings reveal that case-based learning strengthens basic interpretation and application skills by focusing on the clinical context and decision-making steps. This aligns with the approaches supported by case discussions, small group studies, and competency-based modules in the literature, especially in radiology education (5-7).

The significant and permanent gain in the US component may be associated with the enrichment of the training with visual materials as well as early and intensive practices, working in small groups, and using peer-supported or peer-educator models. In particular, abundant visual content, scenarios supported by videos, and applications on live mannequins make the learning process more concrete and permanent. Comprehensive reviews in recent years show that peer educators can achieve a level of proficiency close to that of the instructor in teaching activities. Appropriate preparation, audiovisual materials, and live applications contribute significantly to skill development, and short-session practices with frequent feedback significantly increase learning (8,9).

The modality-specific segregation seen in the retention window is also consistent with the trend in the literature: POCUS/basic US image acquisition and interpretation skills are significantly maintained at mid-term (6-8 weeks), whereas performance in more complex image acquisition/interpretation processes (e.g., cardiac windows, multistep slice analysis) fades more rapidly (10,11). These findings support the higher persistence observed for US and direct radiography in our study. Some studies have demonstrated that superficial vascular structures can be easily visualized via US by all physicians (12). Our training was designed at the basic US/POCUS level, and the evaluation questions were prepared within this framework, which may explain the retention success in US training. However, a similar retention may not be achieved in a training program targeting more advanced imaging windows or complex analyses.

Another implication is that learning should be distributed and reinforced over time. Systematic reviews focused on radiology education show that spaced learning, retrieval practice, and interleaving strategies can provide significant advantages in both short-term and long-term examinations, and that spaced digital education contributes to knowledge growth and behavioral change in health professions (13-15). Future studies should examine learning curves for CT/MRI with longitudinal designs involving intermittent reinforcement and repeated presentation of rare pathologies. To ensure long-term retention in CT and MRI imaging, incorporating periodic reinforcement and simulation-based case discussions is recommended. Integrating CT/MRI interpretation into longitudinal curricula, supported by spaced digital learning and structured feedback, may help sustain diagnostic accuracy and cognitive retention over time.

In this study, 50 intern physicians participated, representing diverse specialty interests including emergency medicine, internal medicine, surgery, and psychiatry. Although participants reported a wide range of specialty preferences, no statistically significant difference was found between these preferences and the learning outcomes in radiology education. However, students planning to specialize in clinically dynamic disciplines such as emergency medicine or internal medicine may have higher

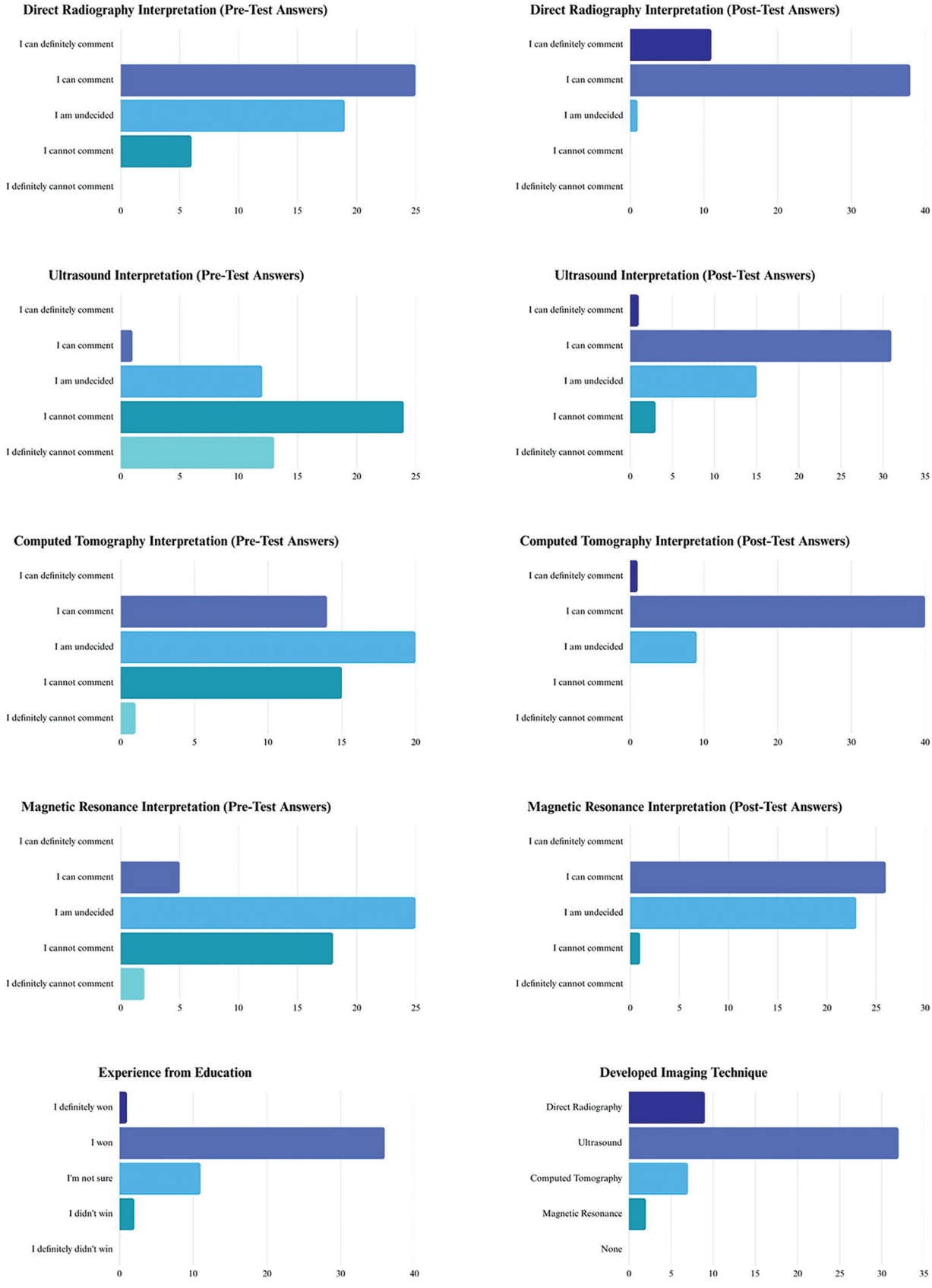


Figure 1. Graphical visualization volunteers' responses

Table 2. Pre- and post-test average score statistics			
Radiological method	**Pre-test**	**Post-test***	**p-value**
Direct radiography (%), median (IQR)	25 (IQR: 25-50)	75 (IQR: 50-75)	**0.001**
Ultrasound (%), median (IQR)	0 (IQR: 0-0)	100 (IQR: 100-100)	**0.001**
Computed tomography (%), median (IQR)	33.3 (IQR: 33.7-66.7)	100 (IQR: 66.7-100)	**0.001**
Magnetic resonance (%), median (IQR)	(IQR: 0-100)	100 (IQR: 0-100)	**0.019**
*: Correct answer percentage, IQR: Interquartile range (25p, 75p)			

Table 3. Pre- and persistence test average score statistics			
Radiological method	**Pre-test**	**Persistence test***	**p-value**
Direct radiography (%), median (IQR)	25 (IQR: 25-50)	75 (IQR: 50-75)	**0.001**
Ultrasound (%), median (IQR)	0 (IQR: 0-0)	85.7 (IQR: 71.4-96.4)	**0.001**
Computed tomography (%), median (IQR)	33.3 (IQR: 33.7-66.7)	77.8 (IQR: 66.7-77.8)	**0.001**
Magnetic resonance (%), median (IQR)	0 (IQR: 0-100)	100 (IQR: 75-100)	**0.001**
*: Correct answer percentage, IQR: Interquartile range (25p, 75p)			

Table 4. Post- and persistence test average score statistics			
Radiological method	**Post-test***	**Persistence test***	**p-value**
Direct radiography (%), median (IQR)	75 (IQR: 50-75)	75 (IQR: 50-75)	0.381
Ultrasound (%), median (IQR)	100 (IQR: 100-100)	85.7 (IQR: 71.4-96.4)	0.059
Computed tomography (%), median (IQR)	100 (IQR: 66.7-100)	77.8 (IQR: 66.7-77.8)	**0.006**
Magnetic resonance (%), median (IQR)	100 (IQR: 0-100)	100 (IQR: 75-100)	**0.001**
*: Correct answer percentage, IQR: Interquartile range (25p, 75p)			

intrinsic motivation and engagement during case-based imaging training, as radiologic interpretation directly contributes to their future clinical decision-making processes.

In the assessment dimension, it is emphasized that standardized and comparable assessment tools are still limited in US education; multiple-choice knowledge tests and self-assessment forms are frequently used. Objective structured clinical exams directed objective procedural skills/objective structured assessment of ultrasound skills (OSAUS) like frameworks that measure practical competence objectively and observationally need more systematic integration (8). In this context, the pre-, post-, and retention-knowledge exams used in our study are effective. It is recommended to add performance-based measures with structured practice stations (e.g., FAST/eFAST OSAUS checklists) at a later stage.

Study Limitations

The main limitations of the study are the single-center design, sample size, and the presence of a subjective perception component in the questionnaire items. In addition, the retention assessment was conducted only one month after the training, which does not allow for definitive conclusions about long-term knowledge retention. Given the partial decline observed in CT and MRI performance, it is possible that this decrease will further progress over longer follow-up periods (e.g., 6 months or 1 year).

Conclusion

The data obtained show that the case-based and practice-oriented mixed curriculum, in the emergency department context, significantly enhances short-term learning and provides high retention in direct radiography and US in one month.

Ethics

Ethics Committee Approval: This study was approved by the institutional review board and ethics committee of Recep Tayyip Erdoğan University in Türkiye (decision no: 2025/357, date: 31.07. 2025).

Informed Consent: Patients' consent was obtained from the patients before starting the study.

Acknowledgment

We want to thank the Department of Emergency Medicine for their hard work and help in data collection.

Footnotes

Authorship Contributions

Surgical and Medical Practices: İ.A., Ö.B., Concept: İ.A., D.H., O.K., Design: M.M.Y., A.B.Ş., H.A., Data Collection or Processing: İ.A., E.G., Analysis or Interpretation: İ.A., Literature Search: İ.A., Writing: İ.A., Ö.B.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Glenn-Cox S, Hird K, Sweetman G, Furness E. Radiology teaching for interns: experiences, current practice and suggestions for improvement. J Med Imaging Radiat Oncol. 2019;63:454-60.
2. Harthoorn FS, Scharenborg SWJ, Brink M, Peters-Bax L, Henssen D. Radiology education for medical students: a qualitative exploration of educational topics, teaching methods and future strategies. BMC Med Educ. 2024;24:891.
3. Umpar S, Sison M. Attitude, skills and clinical competency of radiologic technology interns on general radiography. Psychol Educ Multidiscip J. 2025;34:1077-87.
4. Yazıcı MM, Ataş İ, Güler E, Bilir Ö. Comparison of traditional and online education in airway management. Int J Res Med Sci. 2023;11:1509-13.
5. Fromke EJ, Jordan SG, Awan OA. Case-based learning: Its importance in medical student education. Acad Radiol. 2022;29:1284-86.
6. Sugi MD, Kennedy TA, Shah V, Hartung MP. Bridging the gap: Interactive, case-based learning in radiology education. Abdom Radiol (NY). 2021;46:5503-8.
7. Masthoff M, Pawelka F, Zak G, de Leng B, Darici D, Schindler P, et al. Students' perspective on new teaching concepts for medical studies: case- and competency-based learning in radiology. Insights Imaging. 2025;16:31
8. Recker F, Neubauer R, Dong Y, Gschmack AM, Jenssen C, Möller K, et al. Exploring the dynamics of ultrasound training in medical education: current trends, debates, and approaches to didactics and hands-on learning. BMC Med Educ. 2024;24:1311.
9. Winter L, Neubauer R, Weimer J, Dietrich CF, Wittek A, Schiestl L, et al. Peer teachers as ultrasound instructors? A systematic literature review of peer teaching concepts in undergraduate ultrasound education. BMC Med Educ. 2024;24:1369.
10. Jujo S, Sakka BI, Lee-Jayaram JJ, Kataoka A, Izumo M, Kusunose K, et al. Medical student mid-term skill retention following cardiac point-of-care ultrasound training based on the American Society of Echocardiography curriculum framework. Cardiovasc Ultrasound. 2022;20:26.
11. Matthews L, Contino K, Nussbaum C, Hunter K, Schorr C, Puri N. Skill retention with ultrasound curricula. PLoS One. 2020;15:e0243086.
12. Ataş İ, Yazıcı MM, Hamdioğlu E, Parça N, Kaçan M, Yavaşi Ö, et al. Ultrasonographic evaluation of hypervolemic and normovolemic patients: a comparison of inferior vena cava, subclavian vein, internal jugular vein, and femoral vein diameters and collapsibility indices. Cureus. 2025;17:e77488.
13. Thompson CP, Hughes MA. The effectiveness of spaced learning, interleaving, and retrieval practice in radiology education: a systematic review. J Am Coll Radiol. 2023;20:1092-101.
14. Versteeg M, Hendriks RA, Thomas A, Ommering BWC, Steendijk P. Conceptualising spaced learning in health professions education: a scoping review. Med Educ. 2020;54:205-16
15. Martinengo L, Ng MSP, Ng TR, Ang YI, Jabir AI, Kyaw BM, et al. Spaced digital education for health professionals: Systematic review and meta-analysis. J Med Internet Res. 2024;26:e57760.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.13334
Eurasian J Emerg Med. 2026;25:90-7

Ultrasound Assessment of Quadriceps Muscle Layer Thickness as a Predictor of Outcome in Critically ill Patients

Sham Antony, Saurav Mitra Mustafi, Vandana Talwar, Ajay Kumar, Santvana Kohli

Vardhman Mahavir Medical College and Safdarjung Hospital, Clinic of Critical Care Medicine, New Delhi, İndia

Abstract

Aim: Muscle weakness in critically ill patients is associated with complications, but its accurate quantification at the bedside remains challenging. Our aim was to assess whether loss of quadriceps muscle layer thickness (QMLT) predicts mortality, and to examine its correlation with duration of mechanical ventilation and other severity indices.

Materials and Methods: A total of 100 intensive care unit (ICU) patients were prospectively assessed with QMLT measurements on the dominant side on days 1, 3, and 7 of the ICU stay, or on the day of discharge, whichever occurred earlier.

Results: Percentage decline in QMLT on days 3 and 7 (or discharge) was a significant predictor of mortality [for day 3: hazard ratio (HR): 1.06, 95% confidence interval (CI): 1.03-1.1, p=0.001; for day 7 or discharge: HR: 1.05, 95% CI: 1.02-1.07, p=0.0002]. The decline in QMLT correlated well with mortality (p=0.50, p<0.001 at day 3, and p=0.77, p<0.001 at day 7 or on the day of discharge). A decline in QMLT correlated with duration of mechanical ventilation on day 7 (p=0.46, p<0.001) and with mNUTRIC score on day 3 and on day 7 of ICU stay or on the day of discharge (p=0.36, p <0.001 at day 3; p=0.44, p<0.001 at day 7 or on the day of discharge). The decline in QMLT did not correlate with the duration of ICU stay. A >25.97% decline in QMLT on day 7 of ICU stay was associated with a high area under the curve (AUC) (AUC 0.96; p<0.0001).

Conclusion: Continuous monitoring of QMLT decline by ultrasound independently predicts mortality.

Keywords: ICU patients, ICU outcome, quadriceps muscle, ultrasound

Introduction

Muscle weakness in critically ill patients is linked to a wide range of health challenges (1,2). The causes are multifactorial and include deranged glycaemia, inflammatory stress responses, immobilization, nutrient deficiencies, and various medications—all of which contribute to intensive care unit (ICU)–associated muscle weakness (1,3). Early identification of muscle weakness in critically ill patients is essential (4). Muscle wasting can progress at a rate of up to 2% per day, with the greatest losses observed in patients with multiorgan dysfunction syndrome during the first 10 days of illness (3,5). Such muscle loss can delay recovery and is associated with longer hospital stays, higher mortality, and increased healthcare costs (6).

Accurate bedside quantification of muscle wasting remains a challenge (7,8). Conventional anthropometric measurements are often unreliable in the ICU, while advanced imaging techniques such as computed tomography are impractical for routine use (7,9,10). Ultrasonography (USG) has therefore emerged as a valuable bedside tool—non-invasive, cost-effective, and widely accessible (4,10-13).

Loss of quadriceps muscle-layer thickness (QMLT) has been shown to correlate with in-hospital survival, duration of mechanical ventilation, and other indicators of severity in the ICU (3,7,14). However, data on this topic remain limited in Asian populations (15). Therefore, we designed this study to assess QMLT, measured by USG, as a predictor of mortality in critically ill patients. The secondary objectives were to examine the relationships between

Corresponding Author: Santvana Kohli, Vardhman Mahavir Medical College and Safdarjung Hospital, Clinic of Critical Care Medicine, New Delhi, İndia
E-mail: dr.santvana.kohli@gmail.com **ORCID ID:** orcid.org/0009-0000-4496-1266

Cite this article as: Antony S, Mustafi SM, Talwar V, Kumar A, Kohli S. Ultrasound assessment of quadriceps muscle layer thickness as a predictor of outcome in critically ill patients. Eurasian J Emerg Med. 2026;25:90-7.

Received: 30.09.2025
Accepted: 15.12.2025
Published: 26.01.2026

©Copyright 2026 The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License.

QMLT and the duration of mechanical ventilation, the length of ICU stay, and the modified Nutrition Risk in the Critically Ill (mNUTRIC) score among patients admitted to our ICU.

Materials and Methods

This prospective observational study involved 100 patients admitted to the ICU of our institute and was conducted after obtaining this study received approval the Vardhman Mahavir Medical College and Safdarjung Hospital Institutional Ethics Committee (desicion number: IEC/VMMC/SJH/Thesis/06/2022/CC-47, date: 11.07.2022). Toledo et al. (7) reported hazard ratios (HRs) of 2.1 for requiring mechanical ventilation, 3.7 for probability of ICU survival, and 4.5 for probability of hospital survival, based on thigh muscle thickness. Using these values as references, the minimum required sample size to achieve 80% study power at a 5% significance level was calculated as 97.13 patients. To reduce the margin of error, 100 patients were recruited.

All adults aged 18 years and above who were expected to require mechanical ventilation for at least 48 hours were included after obtaining written informed consent. Patients were excluded if they had documented neuromuscular disease (e.g., myopathy, neuropathy), cerebrovascular accident, lower limb amputation or orthopedic surgery, pregnancy, requirement for prone positioning, transfer from another hospital after an ICU stay of more than 48 hours, extubation within 48 hours, or a hospital stay of less than 72 hours due to death or discharge.

During the first 24 hours after ICU admission, demographic, biochemical, and clinical data were collected. Acute Physiology and Chronic Health Evaluation II (APACHE II) scores, Sequential Organ Failure Assessment (SOFA) scores, and mNUTRIC scores were recorded at admission. Quadriceps muscle thickness was measured using a portable B-mode Sonosite M-Turbo (Digi 600 A Pro) ultrasound device with a 6-12 MHz linear array probe. The assessor was trained through 20 supervised scans to ensure standardization. All measurements were taken in a semi-recumbent position (30°-45°), with the knees extended and the toes pointing upward.

A line was drawn between the anterior superior iliac spine and the upper pole of the patella. The point at the junction between the upper two-thirds and the lower one-third of this line was marked, and ultrasound gel was applied. The transducer was held perpendicular to the line, and the depth was adjusted to visualize the femur. Once an optimal image was obtained, it was frozen. QMLT was measured as the distance between the upper surface of the femoral bone and the lower border of the superficial fascia of the rectus femoris, thereby including both the rectus femoris and vastus intermedius muscles. Each measurement was taken twice on the dominant leg (defined by the hand used for eating), and the average value was recorded. QMLT measurements were obtained on days 1, 3, and 7 of the ICU stay, or earlier if the patient was discharged.

All patients were monitored for biochemical parameters, treatments received, duration of mechanical ventilation, length of ICU stay, total hospital stay, and 28-day mortality. Caloric and protein intakes (calorie/protein debt) from days 1 to 7 were also recorded, using baseline requirements of 25 kcal/kg/day and 1.2 g/kg/day of protein. Intake was documented on days 1, 3, and 7.

Statistical Analysis

Data entry was performed using Microsoft Excel, and statistical analyses were carried out using the Statistical Package for the Social Sciences (SPSS), version 25.0 (IBM, Chicago, USA). Quantitative data with a normal distribution were expressed as mean ± SD, whereas non-normally distributed data were expressed as median (interquartile range). Normality was assessed using the Shapiro-Wilk test. Non-normal data were analyzed using the Mann-Whitney U test, and normally distributed data were analyzed using the independent t-test.

ROC curves were used to determine the cut off point, sensitivity, specificity, positive predictive value, and negative predictive value for predicting mortality based on the percentage decrease in QMLT. Kaplan-Meier survival curves with log-rank testing were used to assess overall survival. Univariate and multivariate Cox proportional hazards regression analyses were performed to identify significant risk factors for mortality. Spearman's rank correlation coefficient was used to evaluate the relationships between the percentage change in QMLT and duration of mechanical ventilation (days), ICU stay (days), and mNUTRIC score. A point-biserial correlation was used to assess the association between the percentage change in QMLT and mortality.

Results

The demographic characteristics of the study population are presented in Table 1. The mean age of participants was 38.1±18 years; 67% were male and 33% were female. The mean body mass index (BMI) was 22.36±3.21 kg/m^2. A majority of the patients (73%) had no comorbidities. Among the remaining 27%, diabetes mellitus was the most common condition (14%), followed by chronic obstructive pulmonary disease (11%). The median APACHE II, SOFA, and mNUTRIC scores at admission were 14, 7, and 2, respectively. Based on a standard daily requirement of 25 kcal/kg and 1.2 g/kg of protein, the median calorie intake increased from 47.6% on day 1 to 77.4% on day 7; the median protein intake increased from 15.9% on day 1 to 49.2% on day 7. All patients required mechanical ventilation; the median

duration was 5 days, and the mean length of ICU stay was 6 days. The overall mortality in the cohort was 51%.

Table 1. Demographic data and patient parameters

Variables	Mean (±SD), median (IQR), n (%)
Age (years)*	38.1 (±18)
Gender#	
Male	67 (67)
Female	33 (33)
Height (cm) #	171.50 (9.56)
Weight (kg) #	66.02 (11.9)
BMI (kg/m^2) #	22.36 (3.21)
Comorbidity#	
Diabetes mellitus	14 (14)
COPD	11 (11)
Hypertension	9 (9)
Renal	3 (3)
Cardiac	2 (2)
Absent	73 (73)
APACHE II**	14 (10.2-17)
SOFA score**	7 (4-9)
mNUTRIC score**	2 (1-3)
Serum glucose (mg/dL)*	141.8 (70.4)
Creatinine (mg/dL) *	2.1 (2.8)
Urea (mg/dL)*	75.7 (70)
Sodium (mEq/L)*	138.8 (8.3)
Potassium (mEq/L)*	4.5 (1)
Calcium (mg/dL)*	7.6 (1.1)
Magnesium (mg/dL)*	2.2 (0.5)
Phosphate (mg/dL)*	4.1 (2)
Procalcitonin (ng/mL) *	22 (20.1)
Lactate (mmol/L)*	2.5 (2.3)
Bicarbonate (mmol/L)*	22.7 (6.9)
pH*	7.355 (0.1)
Percentage of expected calorie intake#	
Day 1	47.6 (29.7)
Day 3	74.4 (27.6)
Day 7	77.4 (36.1)
Percentage of expected protein intake#	
Day 1	15.9 (21)
Day 3	44.55 (32.4)
Day 7	49.25 (36.6)
Mechanical ventilation#	100 (100)
Duration of mechanical ventilation (days)**	5 (4-7)
Length of ICU stay (days)**	6 (5-7)
Mortality#	51 (51)

*: Mean (±SD), **: Median (IQR), #: n (%), APACHE II: Acute Physiology and Chronic Health Evaluation 2, SOFA: Sequential Organ Failure Assessment, mNUTRIC: Modified Nutritional Risk in Critically Ill, ICU: Intensive care unit, COPD: Chronic obstructive pulmonary disease, SD: Standard deviation, IQR: Interquartile range

There were no significant differences in age or BMI between survivors and non-survivors. However, non-survivors had significantly higher APACHE II, SOFA, and mNUTRIC scores at admission compared with survivors [APACHE II: confidence interval (CI): 0.63-5.1, p=0.013; SOFA: CI: 1.42-3.81, p<0.0001; mNUTRIC: Z=3.9, p<0.0001]. Median urea and creatinine levels were also significantly higher in non-survivors (Z=3.61, p=0.0003; Z=3.85, p=0.0001, respectively). Non-survivors were more acidotic, with significantly lower mean pH (CI: -0.25 to 5.2; p=0.009). The median duration of mechanical ventilation was significantly lower in survivors (Z=3.29; p=0.001), although the length of ICU stay was comparable between groups.

Median calorie and protein intake on day 1 did not differ significantly between survivors and non-survivors. However, non-survivors demonstrated significantly lower calorie intake (day 3: Z=3.42, p=0.0006; day 7: Z=3.10, p=0.002) and protein intake (day 3: Z=3.12, p = 0.002; day 7: Z=2.81, p=0.005). A significant reduction in QMLT in the dominant leg was observed on ICU days 3 and 7 (Z=4.85, p < 0.0001; Z=7.86, p<0.0001, respectively). These results are summarized in Table 2.

The percentage decline in QMLT was evaluated on days 3 and 7 to determine its predictive value for mortality. ROC curve analysis identified cut-offs of a 14.5% decline on day 3 and a 25.97% decline on day 7 (Table 3, Figure 1). A decline exceeding these thresholds had a specificity of 85.7% and a sensitivity of 62.8% for predicting mortality on day 3. By day 7, sensitivity and specificity increased to 86.3% and 93.9%, respectively. The positive predictive value improved from 82.1% on day 3 to 93.6% on day 7, and the negative predictive value increased from 68.9% to 86.8%. The diagnostic accuracy increased from 72% on day 3 to 90% on day 7. The area under the curve was significantly higher for the day-7 cut off (0.96) compared with the day-3 cut-off (0.75) (p<0.0001).

Cox regression analyses were performed using age, BMI, APACHE II, SOFA, and mNUTRIC scores, along with percentage decline in QMLT, in two models: day 3 (model A) and day 7 (model B) (Table 4). The Percentage decline in QMLT on days 3 and 7 remained the only significant predictor of mortality: HR: 1.06 (95% CI: 1.03-1.10), p=0.001 for day 3; HR: 1.05 (95% CI: 1.02-1.07), p=0.0002 for day 7 This suggests that each percentage decline in QMLT is associated with an approximately 5-6% increase in mortality risk during the ICU stay.

Kaplan-Meier difference in survival curves showed a significant between patients with QMLT declines below versus above the identified thresholds on day 3 and on day 7 or on the day of discharge (log-rank test p=0.001 and p<0.0001, respectively). Patients with smaller declines demonstrated consistently better survival (Figure 2).

Table 2. Comparison of patient parameters between survivors and non-survivors

Parameters	Survivors	Non survivors	Z/mean diff (CI)/χ^2 value	p-value
Age (years)	35.35±17.25	40.57±20.13	5.22 (2.23-12.67)	0.17*
Body mass index (kg/m²)	22.47±3.11	22.25±3.33	0.22 (-1.06-1.5)	0.74*
APACHE II score	12.69±4.62	15.55±6.43	2.86 (0.63-5.1)	**0.013***
SOFA score	5.57±2.38	8.19±3.51	2.62 (1.42-3.81)	**<0.0001***
mNUTRIC score	1 (1-2)	3 (1.5-4)	3.9	<0.0001†
Creatinine (mg %)	0.8 (0.5-1.2)	1.5 (0.85-3.25)	3.85	0.0001†
Urea (mg/dL)	37 (21-57)	71 (35.5-137)	3.61	0.0003†
Procalcitonin (ng/mL)	12.3 (2.5-40.9)	15.7 (3.65-50)	0.36	0.72†
Lactate (mmol/L)	1.4(1-2.6)	1.9 (1.15-3.9)	1.81	0.07†
Bicarbonate (mmol/L)	23.98±7.38	21.5±6.33	2.48 (-0.25-5.2)	0.08*
pH	7.38±0.09	7.33±0.12	0.06 (0.01-0.1)	**0.009***
Duration of mechanical ventilation (days)	5 (3-6)	6 (5-7)	3.29	0.001†
Length of ICU stay (days)	6 (5-7)	6 (5-7.5)	0.47	0.64†
% calorie intake day 1	59.21 (16.05-75.32)	43.8 (17.93-66.58)	1.06	0.29†
% calorie intake day 3	84 (72.91-97.05)	67 (49.44-81.922)	3.42	**0.0006†**
% calorie intake day 7	95.88 (91.865-113.75)	67.5 (34.77-90)	3.1	**0.002†**
% protein intake day 1	0 (0-35.71)	0 (0-33.69)	1	0.32†
% protein intake day 3	60.48 (28.48-75.94)	34 (13.815-52.12)	3.12	**0.002†**
% protein intake day 7	81.65 (36.88-87.575)	29.49 (19.9-56.96)	2.81	**0.005†**
% decline in QMLT day 3	9.5 (4.808-12.632)	16.73 (11.655-25.516)	4.85	**<0.0001†**
% decline in QMLT day 7 or discharge	12.38 (8.466-15.819)	34.45 (28.954-40.017)	7.96	**<0.0001†**

*: Independent t test, †: Mann-Whitney test
QMLT: Quadriceps muscle layer thickness, APACHE II: Acute Physiology and Chronic Health Evaluation 2, SOFA: Sequential Organ Failure Assessment, mNUTRIC: Modified Nutritional Risk in Critically Ill, COPD: Chronic obstructive pulmonary disease, CI: Confidence interval

Table 3. Sensitivity, specificity, positive predictive value and negative predictive value of percentage decrease in QMLT for predicting mortality

Variables	Percentage decline in QMLT dominant leg day 3	Percentage decline in QMLT dominant leg day 7 or discharge
Area under the ROC curve (AUC)	0.78	0.96
Standard error	0.047	0.016
95% confidence interval	0.69-0.86	0.9-0.99
p-value	<0.0001	<0.0001
Cut off	>14.48	>25.97
Sensitivity	62.8	86.3
Specificity	85.7	93.9
PPV	82.1	93.6
NPV	68.9	86.8
Diagnostic accuracy	72	89

QMLT: Quadriceps muscle layer thickness, PPV: Positive predictive value, NPV: Negative predictive value, AUC: Area under the curve

A significant correlation was found between a decline in QMLT and duration of mechanical ventilation on day 7 (p=0.46, p<0.001), and between a decline in QMLT and mNUTRIC scores on days 3 and 7 (p=0.36, p<0.001 and p=0.44, p < 0.001, respectively). A decline in QMLT did not correlate significantly with ICU length of stay (Table 5), suggesting limited reliability for predicting ICU length of stay. The decline in QMLT correlated strongly with mortality (p=0.50, p<0.001 on day 3; p=0.77, p<0.001 on day 7).

Discussion

This study reflects the characteristic demographics of the Asian subcontinent, with all patients in our cohort requiring mechanical ventilation. The high median SOFA scores highlight the severity of illness and explain the elevated mortality rate. This makes our cohort well suited to draw meaningful conclusions regarding QMLT decline as assessed by ultrasound. Although several ultrasound protocols for monitoring muscle mass exist in the literature, we chose to measure quadriceps muscle thickness

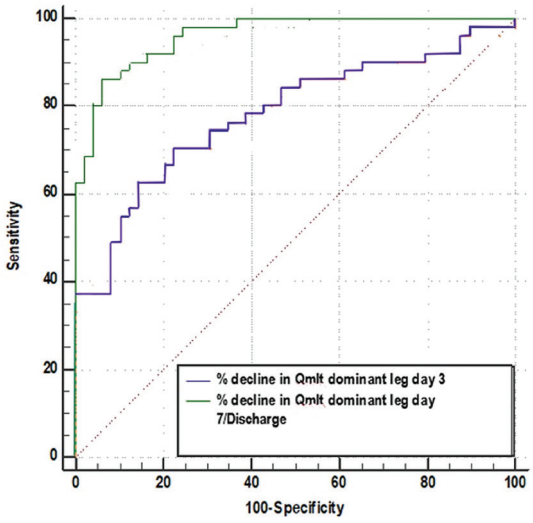


Figure 1. Receiver operating characteristic curve of percentage decrease in QMLT for predicting mortality

QMLT: Quadriceps muscle layer thickness

without compression rather than assessing cross-sectional area (CSA), owing to its simplicity. Previous research has demonstrated comparable results between measurements taken with and without compression, supporting the reliability of this method (4,5,7,10). However, a standardized ultrasound protocol for QMLT assessment has yet to be established (11).

We found that a 24% decrease in QMLT from baseline—measured on day 7 of the ICU stay or on the day of discharge—is a significant risk factor for mortality. A decline in QMLT was also correlated with higher mNUTRIC scores on days 3 and 7 and with prolonged mechanical ventilation by day 7. The association between QMLT decline and mortality is consistent with findings from other studies, which emphasize that serial decline, rather than single-time-point QMLT values, is more clinically meaningful (7).

Furthermore, patients who exhibited a greater loss of QMLT had a lower probability of survival, as demonstrated by Kaplan-Meier analysis. Cox regression analysis identified QMLT decline as an independent risk factor for mortality. Our findings suggest that each percentage-point decline in QMLT corresponds to a 5-6% increase in mortality risk at both time points evaluated. We analyzed decline in QMLT on days 3 and 7 using separate Cox models to avoid confounding between the two time points. Prior studies employing similar statistical methods have shown comparable trends (7,16-18). To the best of our knowledge, however, no previous study has demonstrated QMLT decline as an independent predictor of mortality using Cox proportional hazards analysis. Studies using CT imaging to assess muscle loss in critically ill patients have similarly reported an association between muscle loss and increased mortality (19).

A few studies in the literature, however, report findings that differ from ours (10,20,21). The VALIDUM study, a prospective multicenter trial, concluded that QMLT and CSA assessments were insufficient to identify patients with low muscle mass; patients with a poor prognosis who were unlikely to survive were excluded, and this exclusion may have influenced the results (10). Palakshappa et al. (20) reported a moderate correlation between the decline in rectus femoris CSA and muscle strength, but their study included only 29 predominantly septic patients. A systematic review by Bunnell et al. (21) questioned the utility of neuromuscular ultrasound in assessing critical illness neuromyopathy due to methodological inconsistencies, variable tissue water content, interobserver variability, differences in probe angle, and variation in transducer frequency.

The American Society for Parenteral and Enteral Nutrition recommends screening tools such as the mNUTRIC score for assessing nutritional risk (22). Özdemir et al. (18) argued that mNUTRIC scores may not fully reflect nutritional status because

Table 4. Univariate and multivariate cox proportional hazard regression for significant risk factors of mortality

Variables	Univariate hazards ratio (95% CI)	p-value	Multivariate hazards ratio (95% CI)	p-value
Age (years) A, B	1.01 (0.99-1.03)	0.06	...	...
Body mass index (kg/m²) A, B	0.96 (0.87-1.05)	0.38	...	...
APACHE II score A,B	1.04 (0.99-1.09)	0.15	...	...
SOFA score A, B	1.1 (1.01-1.2)	0.03	...	...
mNUTRIC score A, B	1.28 (1.07-1.53)	0.006	...	...
Percentage decline in QMLT dominant leg day 3 A, B	1.07 (1.04-1.11)	<0.0001	1.06 (1.03-1.1)	**0.001**
Percentage decline in QMLT dominant leg day 7 or discharge A, B	1.051 (1.03-1.07)	<0.0001	1.05 (1.02-1.07)	**0.0002**

Model A: Age, BMI, APACHE II, SOFA, mNUTRIC, QMLT decline percent day 3 dominant side
Model B: Age, BMI, APACHE II, SOFA, mNUTRIC, QMLT decline percent day 7 dominant side
QMLT: Quadriceps muscle layer thickness, APACHE II: Acute Physiology and Chronic Health Evaluation 2, SOFA: Sequential Organ Failure Assessment, mNUTRIC: Modified Nutritional Risk in Critically Ill, COPD- Chronic Obstructive Pulmonary Disease, CI: Confidence interval

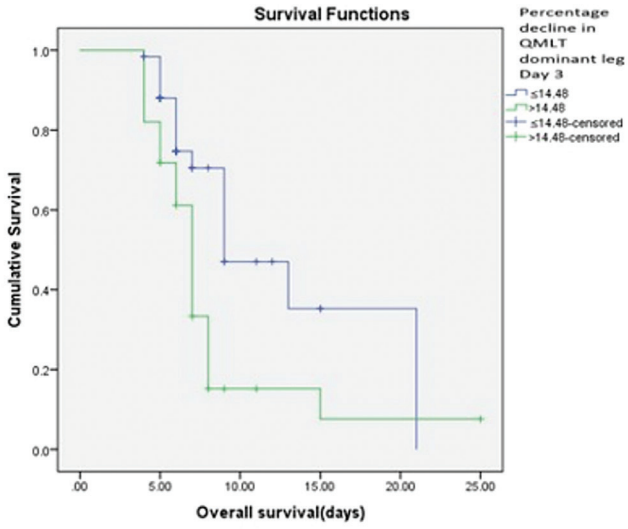


Figure 2.1. Kaplan Meier survival analysis curve to compare overall survival between percentage decline in QMLT in dominant leg on day 3 (log rank test, p=0.001)

QMLT: Quadriceps muscle layer thickness

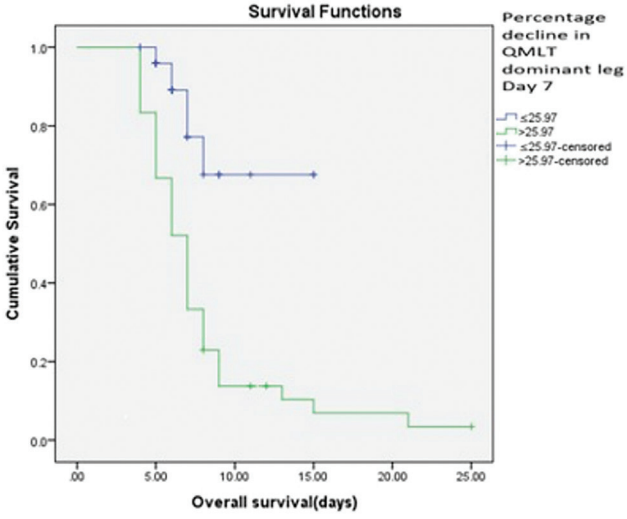


Figure 2.2. Kaplan Meier survival analysis curve to compare overall survival between percentage decline in QMLT in dominant leg on day 7 or discharge (log rank test, p<0.0001)

QMLT: Quadriceps muscle layer thickness

they lack anthropometric components. This highlights the need for improved nutritional risk assessment tools for ICU patients. The moderate correlation between QMLT decline and mNUTRIC scores on days 3 and 7 in our study suggests that QMLT may serve as a useful adjunctive tool for nutritional assessment. Previous research indicates that muscle loss is greatest during the first week of mechanical ventilation (5,7). Our study also demonstrated a moderate correlation between QMLT decline and duration of mechanical ventilation, although no correlation was observed with ICU length of stay. These findings support the potential role of QMLT decline as a prognostic marker.

Non-survivors in our cohort consistently failed to meet the recommended calorie and protein targets from day 2 onward, compared with survivors. Some studies have shown that achieving individualized caloric and protein targets in critically ill patients may reduce mortality by up to 50% (23). However, another study found that this benefit was limited to female patients, likely due to their lower baseline muscle mass (24). Such mixed evidence underscores the need for more robust and objective parameters—such as USG-based QMLT—provided they can be standardized. These may ultimately outperform traditional nutritional targets and risk scores.

Nutritional status plays a crucial role in preserving skeletal muscle mass, and inadequate calorie–protein intake accelerates muscle catabolism in critically ill patients. Studies have shown that ICU patients can lose up to 2% of muscle mass per day during the first week, and undernutrition amplifies this decline. Poor nutritional intake reduces amino acid availability for muscle protein synthesis, leading to a measurable decrease in QMLT. Research demonstrates that greater early muscle loss, including QMLT reduction, is strongly associated with prolonged mechanical ventilation and higher mortality rates. Undernutrition heightens the inflammatory and stress responses, further driving catabolism and worsening muscle wasting.

Study Limitations

Our study has certain limitations. Sepsis- and trauma-related fluid shifts may have been because intramuscular fluid accumulation could have affected QMLT measurements. Muscle edema can mask true muscle atrophy, resulting in measurements that appear deceptively stable or even elevated. This complicates

Table 5. Correlation of QMLT with various parameters

Percentage change in QMLT	Duration of mechanical ventilation (days)	Duration of ICU stay (days)	mNUTRIC score	Mortality
Dominant leg day 3	p* 0.18, p=0.08	p* -0.04, p 0.8	p* 0.36, **p<0.001**	p** 0.50, **p<0.001**
Dominant leg day 7	p* 0.46, **p<0.001**	p* 0.17, p 0.09	p* 0.44, **p<0.001**	p** 0.77, **p<0.001**

*: Spearman's rank correlation,**: Point-biserial correlation, mNUTRIC: Modified Nutritional Risk in Critically Ill, ICU: Intensive care unit, QMLT: Quadriceps muscle layer thickness

the interpretation of our measurements. Second, this single-center observational study may therefore serve primarily as a hypothesis-generating investigation. Additionally, assessing quadriceps dimensions in a single plane may not accurately represent muscle loss throughout the body. This may explain why we observed only a moderate correlation with the duration of mechanical ventilation and no correlation with the ICU length of stay. The quadriceps muscle contains both type I and type II fibers—type II fibers serving as power generators and type I fibers as stabilizers (4). Our measurement technique did not allow differentiation between fiber types, which may also influence the interpretation of muscle loss patterns.

Conclusion

Continuous monitoring of QMLT by USG shows that loss of muscle mass during ICU stay is an independent predictor of mortality and correlates with mNUTRIC score and increased days of mechanical ventilation by day 7. This tool is applicable at the bedside, noninvasive, and cost-effective. Intensivists should thus prioritize the routine evaluation of QMLT and adjust nutritional support accordingly. These findings underscore the importance of integrating QMLT assessment into clinical practice to improve patient care in intensive care settings.

Highlights

Continuous monitoring of quadriceps muscle layer thickness using ultrasound is a cost-effective, safe, and valuable tool for intensivists. It can be used to monitor recovery as well as to predict mortality.

Ethics

Ethics Committee Approval: this study received approval the Vardhman Mahavir Medical College and Safdarjung Hospital Institutional Ethics Committee (desicion number: IEC/VMMC/SJH/Thesis/06/2022/CC-47, date: 11.07.2022).

Informed Consent: All adults aged 18 years and above who were expected to require mechanical ventilation for at least 48 hours were included after obtaining written informed consent.

Footnotes

Authorship Contributions

Surgical and Medical Practices: S.M.M., V.T., A.K., Concept: S.A., V.T., Design: S.A., V.T., Data Collection or Processing: S.A., S.M.M., Analysis or Interpretation: S.M.M., A.K., S.K., Literature Search: S.A., V.T., S.K., Writing: S.M.M., A.K., S.K.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Fan E, Cheek F, Chlan L, Gosselink R, Hart N, Herridge MS, et al. An official American Thoracic Society Clinical Practice guideline: the diagnosis of intensive care unit-acquired weakness in adults. Am J Respir Crit Care Med. 2014;190:1437-46.
2. Katari Y, Srinivasan R, Arvind P, Hiremathada S. Point-of-care ultrasound to evaluate thickness of rectus femoris, vastus intermedius muscle, and fat as an indicator of muscle and fat wasting in critically ill patients in a multidisciplinary intensive care unit. Indian J Crit Care Med. 2018;22:781-8.
3. Rajagopal K, Vijayan D, Thomas SM. Association of SOFA score with severity of muscle wasting in critically ill patients: a prospective observational study. Indian J Crit Care Med. 2023;27:743-7.
4. Parry SM, El-Ansary D, Cartwright MS, Sarwal A, Berney S, Koopman R, et al. Ultrasonography in the intensive care setting can be used to detect changes in the quality and quantity of muscle and is related to muscle strength and function. J Crit Care. 2015;30:1151.
5. Puthucheary ZA, Rawal J, McPhail M, Connolly B, Ratnayake G, Chan P, et al. Acute skeletal muscle wasting in critical illness. JAMA. 2013;310:1591-600.
6. Pardo E, El Behi H, Boizeau P, Verdonk F, Alberti C, Lescot T. Reliability of ultrasound measurements of quadriceps muscle thickness in critically ill patients. BMC Anesthesiol. 2018;18:205.
7. Toledo DO, Freitas BJ, Dib R, Pfeilsticker FJDA, Santos DMD, Gomes BC, et al. Peripheral muscular ultrasound as outcome assessment tool in critically ill patients on mechanical ventilation: An observational cohort study. Clin Nutr ESPEN. 2021;43:408-14.
8. Jolley SE, Bunnell AE, Hough CL. ICU-Acquired weakness. Chest. 2016;150:1129-40.
9. Toledo DO, Carvalho AM, Oliveira AMRR, Toloi JM, Silva AC, Francisco de Mattos Farah J, et al. The use of computed tomography images as a prognostic marker in critically ill cancer patients. Clin Nutr ESPEN. 2018;25:114-20.
10. Paris MT, Mourtzakis M, Day A, Leung R, Watharkar S, Kozar R, et al. Validation of bedside ultrasound of muscle layer thickness of the quadriceps in the critically ill patient (VALIDUM study). JPEN J Parenter Enteral Nutr. 2017;41:171-80.
11. Weinel LM, Summers MJ, Chapple LA. Ultrasonography to measure quadriceps muscle in critically ill patients: a literature review of reported methodologies. Anaesth Intensive Care. 2019;47:423-34.
12. Galindo Martín CA, Ubeda Zelaya RDC, Monares Zepeda E, Lescas Méndez OA. ROUNDS studies: relation of OUtcomes with nutrition despite severity-round one: ultrasound muscle measurements in critically ill adult patients. J Nutr Metab. 2018;2018:7142325.
13. Zhang W, Wu J, Gu Q, Gu Y, Zhao Y, Ge X, et al. Changes in muscle ultrasound for the diagnosis of intensive care unit acquired weakness in critically ill patients. Sci Rep. 2021;11:18280.
14. Formenti P, Umbrello M, Coppola S, Froio S, Chiumello D. Clinical review: peripheral muscular ultrasound in the ICU. Ann Intensive Care. 2019;9:57.
15. Sharon T, Nayak SG, Shanbhag V, Hebbar S. An observational study of nutritional assessment, prescription, practices, and its outcome among critically ill patients admitted to an intensive care unit. Indian J Crit Care Med. 2024;28:364-8.
16. Hadda V, Kumar R, Khilnani GC, Kalaivani M, Madan K, Tiwari P, et al. Trends of loss of peripheral muscle thickness on ultrasonography and its relationship with outcomes among patients with sepsis. J Intensive Care. 2018;6:81.
17. Weijs PJ, Looijaard WG, Dekker IM, Stapel SN, Girbes AR, Oudemans-van Straaten HM, et al. Low skeletal muscle area is a risk factor for mortality in mechanically ventilated critically ill patients. Crit Care. 2014;18:R12.

18. Özdemir U, Özdemir M, Aygencel G, Kaya B, Türkoğlu M. The role of maximum compressed thickness of the quadriceps femoris muscle measured by ultrasonography in assessing nutritional risk in critically-ill patients with different volume statuses. Rev Assoc Med Bras (1992). 2019;65:952-8.
19. Moisey LL, Mourtzakis M, Cotton BA, Premji T, Heyland DK, Wade CE, et al. Skeletal muscle predicts ventilator-free days, ICU-free days, and mortality in elderly ICU patients. Crit Care. 2013;17:R206.
20. Palakshappa JA, Reilly JP, Schweickert WD, Anderson BJ, Khoury V, Shashaty MG, et al. Quantitative peripheral muscle ultrasound in sepsis: muscle area superior to thickness. J Crit Care. 2018;47:324-30.
21. Bunnell A, Ney J, Gellhorn A, Hough CL. Quantitative neuromuscular ultrasound in intensive care unit-acquired weakness: a systematic review. Muscle Nerve. 2015;52:701-8.
22. McClave SA, Taylor BE, Martindale RG, Warren MM, Johnson DR, Braunschweig C, et al. Guidelines for the provision and assessment of nutrition support therapy in the adult critically ill patient: society of critical care medicine (SCCM) and American Society for Parenteral and Enteral Nutrition (A.S.P.E.N.). JPEN J Parenter Enteral Nutr. 2016;40:159-211.
23. Weijs PJ, Stapel SN, de Groot SD, Driessen RH, de Jong E, Girbes AR, et al. Optimal protein and energy nutrition decreases mortality in mechanically ventilated, critically ill patients: a prospective observational cohort study. JPEN J Parenter Enteral Nutr. 2012;36:60-8.
24. Strack van Schijndel RJ, Weijs PJ, Koopmans RH, Sauerwein HP, Beishuizen A, Girbes AR. Optimal nutrition during the period of mechanical ventilation decreases mortality in critically ill, long-term acute female patients: a prospective observational cohort study. Crit Care. 2009;13:R132.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.46338
Eurasian J Emerg Med. 2026;25: 98-110

First Aid Knowledge and Attitudes of Restaurant Employees in Foreign Body Aspiration Accidents: A District-Based Study in İstanbul

Mahmut Talha Uçar[1], Fatih Yurt[2], Esra Çiçek[3], Gül Amcaoğlu[2], Hümeyra Kondakçı[2], Merve Nur Ülker[2], Güven Bektemür[1]

[1]University of Health Scienes Türkiye Hamidiye Faculty of Medicine, Department of Public Health, İstanbul, Türkiye
[2]University of Health Sciences Türkiye Hamidiye Faculty of Medicine, İstanbul, Türkiye
[3]Altınekin District State Hospital, Clinic of Emergency Medicine, Konya, Türkiye

Abstract

Aim: The aim of this study is to examine the level of first aid knowledge and attitudes of the personnel working in restaurant establishments in a selected district of İstanbul province.

Materials and Methods: The sample of the study consists of 44 restaurant/restaurant employees. SPSS 22 Package Program was used to analyze the data.

Results: Among participants, 31.8% had first aid knowledge-skills, 15.3% had encountered a tracheal foreign body incident, and 6.9% had intervened. Knowledge-skills specific to tracheal foreign body were present in 28.4%, and 19.3% held a first aid certificate; 59.6% were willing to receive training. Knowledge scores differed significantly by first aid knowledge-skills and by encountering or intervening in such cases. Attitude scores differed significantly by age, workplace position, years of employment, first aid knowledge-skills, prior intervention, and certification status.

Conclusion: The rates of having first aid knowledge, skills, and prior intervention experience among restaurant employees were low despite their frontline role in settings where choking incidents are likely to occur. The strong association between higher knowledge and attitude scores and having first aid training or experience indicates that structured education can effectively improve preparedness. Given that most participants were willing to receive training, incorporating regular first aid and basic life support programs into workplace safety policies for the food service sector could strengthen bystander response and reduce preventable deaths from foreign body airway obstruction.

Keywords: Airway obstruction, first aid, foreign body aspiration, respiratory aspiration, restaurants

Introduction

Foreign body aspiration (FBA) is a time-critical medical emergency in which delays in recognition or intervention may rapidly lead to hypoxia and death (1). Survival often depends on the ability of bystanders to deliver effective first aid, most notably abdominal thrusts (the heimlich maneuver), which remain the primary method for relieving airway obstruction in conscious individuals (2). Although national data on choking events in restaurant environments in Türkiye are lacking, foreign body airway obstruction (FBAO) carries an adult fatality rate exceeding 3% globally (3), and is accurately identified in fewer than 10% of cases requiring assistance (4).

Choking episodes frequently occur during meals, making public dining environments a high-risk setting. Recent epidemiological evidence from Türkiye, based on 192 media-reported cases, indicates that 21.4% of FBA incidents took place in restaurants, with food items predominating as aspirated materials (5). This pattern highlights the need to assess and strengthen the preparedness of restaurant personnel, who are often the only individuals present during the critical early moments of airway obstruction.

Consistently, international studies report food as the leading cause of FBA across all age groups (6,7), yet the readiness of non-medical frontline workers in dining settings remains insufficiently

Corresponding Author: Mahmut Talha Uçar MD, University of Health Scienes Türkiye Hamidiye Faculty of Medicine, Department of Public Health, İstanbul, Türkiye
E-mail: drtalhaucar@gmail.com **ORCID ID:** orcid.org/0000-0002-1433-4832

Cite this article as: Uçar MT, Yurt F, Çiçek E, Amcaoğlu G, Kondakçı H, Ülker MN, et al. First aid knowledge and attitudes of restaurant employees in foreign body aspiration accidents: a district-based study in İstanbul. Eurasian J Emerg Med. 2026;25: 98-110.

Received: 17.07.2025
Accepted: 15.12.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

studied. Despite their pivotal role in recognizing and responding to choking, restaurant workers are not included in mandatory first aid certification frameworks in Türkiye. This gap underscores the need for evidence to guide workplace safety regulations and community-based emergency response strategies.

This study therefore aims to evaluate the first aid knowledge and attitudes of restaurant workers in a district of İstanbul, providing baseline data to inform targeted training programs and support policy development to enhance emergency response capacity in restaurant settings.

Materials and Methods

Study Design

This study was designed as a cross-sectional study. The research hypotheses are as follows:

H1-1: The sociodemographic characteristics of restaurant staff substantially affect their knowledge of first aid for FBA.

H1-0: The sociodemographic characteristics of restaurant staff do not substantially affect their knowledge of first aid for FBA.

H2-1: The first aid training and experience of restaurant staff significantly affect their knowledge of first aid for FBA.

H2-0: The first aid training and experience of restaurant staff do not significantly affect their knowledge of first aid for FBA.

H3-1: Restaurant staff who have received first aid training and have experience in FBA incidents have significantly more positive attitudes toward first aid interventions.

H3-0: There is no significant relationship between first aid training, experience, and attitudes toward first aid interventions.

Study Population and Sample

The study population comprises employees from 260 restaurants situated in the Üsküdar area. The number and locations of these restaurants were acquired from the Licensing Department of the Üsküdar Municipality.

The district consists of 33 neighborhoods, with the distribution of restaurants as follows: Twenty-two neighborhoods contain between 1 and 10 restaurants, five neighborhoods contain between 11 and 20, four neighborhoods contain 21 or more, and two neighborhoods have no restaurants at all. A total of 44 restaurants were included in the sample. One restaurant was selected from each of the 22 neighborhoods with 1-10 restaurants, two restaurants from each of the five neighborhoods with 11-20 restaurants, and three restaurants from each of the four neighborhoods with 21 or more restaurants.

A simple random sampling technique was used for sample selection. Restaurants were sorted according to their license numbers and neighborhoods, and 44 establishments were randomly chosen. Furthermore, 44 other restaurants were randomized as contingency possibilities should a selected restaurant refuse participation.

The study was conducted between March 8 and March 20, 2023.

Every member of staff present during the in-person visits to the selected restaurants were invited to participate in the study. Participation was optional, and no staff members dropped out. The sample comprises all on-site staff who provided consent to participate during the data collecting period. A flow diagram titled “Participant Recruitment and Inclusion Process” has been inserted in the Methods section (Figure 1).

Sample Size and Power Analysis

A priori power analysis was performed with G*Power (version 3.1.9.4) to ascertain the minimal sample size necessary for the investigation. With a medium effect size (Cohen’s d=0.5), a significance threshold of α=0.05, and a target power of 0.80 for a two-tailed independent samples t-test, the necessary total sample size was determined to be 128 individuals.

A priori power analysis was performed with G*Power to ascertain the minimal sample size necessary for the investigation. With a medium effect size (Cohen’s d=0.5), a significance threshold of α=0.05, and a target power of 0.80 for a two-tailed independent samples t-test, the necessary total sample size was determined to be 128 individuals. A total of 176 individuals were recruited for the research, as all eligible and consenting employees present in the visited restaurants during data collection were included.

Data Collection Tools

In this study, staff from restaurants were instructed to independently complete a three-part questionnaire under observation. The questionnaire comprised:

Section 1: 12 questions assessing basic demographic and professional information.

Section 2: 26 questions measuring knowledge levels regarding FBA accidents.

Section 3: 13 questions evaluating attitudes towards FBA accidents.

The researchers constructed the questionnaire items based on the content of the Turkish Ministry of Health First Aid Handbook. Prior to the preparation of the original draft, the questionnaire underwent evaluation by a panel including a public health professor, a public health expert, an emergency medicine

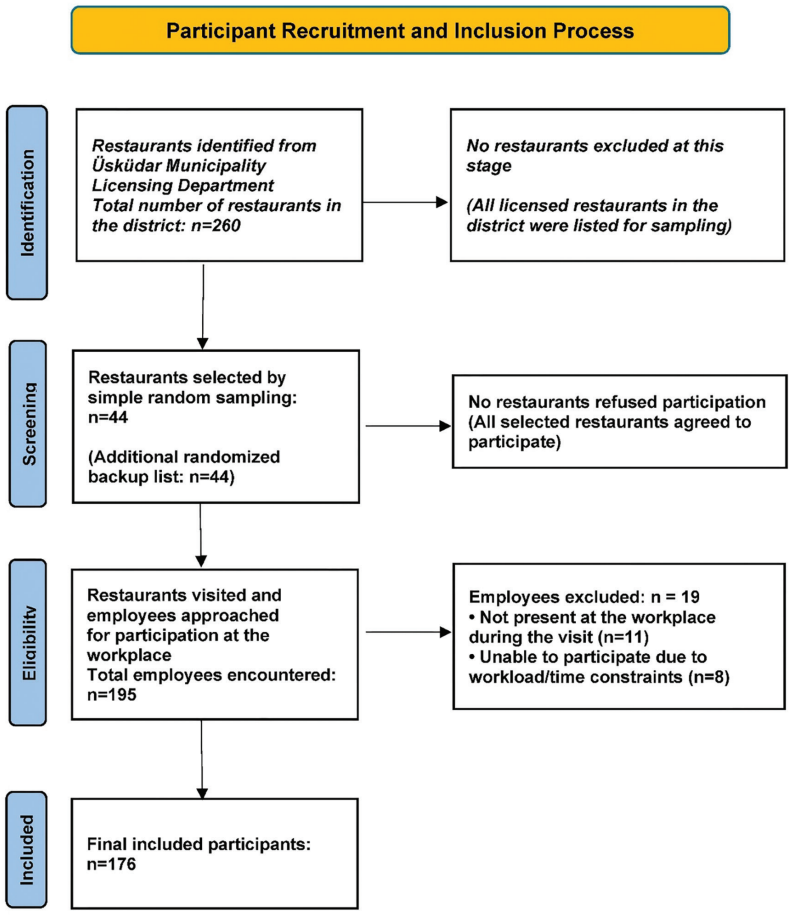


Figure 1. Flow diagram of the participant recruitment and inclusion process

professor, and a psychiatric counselor. Modifications were implemented in response to their suggestions to enhance clarity, content validity, and relevance. A pilot test was administered to evaluate comprehensibility with a cohort of 10 volunteer restaurant staff. Their feedback was utilized to make final modifications prior to deployment. The pilot group data were exclusively utilized for questionnaire refinement and excluded from the final analysis.

The researchers formulated the questions utilizing the Turkish Ministry of Health First Aid Handbook (8). In the knowledge evaluation survey, replies were evaluated as follows: accurate answers earned 1 point, but erroneous answers or "I don't know" responses received 0 points. The total potential scores varied from a minimum of 0 to a maximum of 26.

Knowledge levels were categorized as follows:

Inadequate knowledge level: 0-7 points (<25% correct answers)

Moderate knowledge level: 8-13 points (25-50% correct answers)

Adequate knowledge level: 14-26 points (>50% correct answers)

For the attitude survey, responses were rated on a five-point likert scale: "Strongly agree," "Agree," "Neutral," "Disagree," and "Strongly disagree." The questionnaire included 24 statements, each scored between 1 and 5 points, with a total possible score ranging from a minimum of 13 to a maximum of 65.

The internal consistency of the questionnaire was evaluated using Cronbach's alpha. The reliability analysis demonstrated satisfactory internal consistency for both sections of the instrument, with α=0.789 for the 26-item knowledge section and α=0.777 for the 13-item attitude section.

Ethical Considerations

Before the initiation of the investigation, formal consent was obtained from the University of Health Sciences Hamidiye Scientific Research Ethics Committee (decision no.: 2023/3, date: 10.02.2023). The principles of the Declaration of Helsinki were rigorously followed throughout the study.

Participant Consent

Informed consent was obtained from all participants prior to their inclusion in the study.

Statistical Analysis

The data acquired in this study were analyzed utilizing SPSS version 23. Descriptive statistics were reported as mean ± standard deviation, median (1st quartile-3rd quartile), frequency, and percentage. The normality of continuous variables was evaluated using Shapiro-Wilk tests. Both the knowledge and attitude score distributions deviated from normality; therefore, non-parametric tests were used in the analyses. Consequently, non-parametric tests were utilized. The Mann-Whitney U test was used for pairwise comparisons of non-normally distributed continuous variables, whilst the Kruskal-Wallis test was applied for comparisons involving more than two independent groups. Upon discovering significant differences, post hoc pairwise comparisons were performed with the Dunn-Bonferroni test. A p value of less than 0.05 was deemed statistically significant.

Results

Of the participants, 79% were male, and 42% belonged to the 17-30 age demographic. Regarding educational attainment, 55.7% were high school graduates. Concerning employment positions, 39.8% were employed as waiters. The distribution of work experience indicated that 50.6% had been working for 0 to 8 years. Furthermore, 36.9% of participants were employed in firms with 1 to 14 employees. 31.8% of participants reported possessing knowledge and abilities in first aid procedures. The incidence of those who had previously experienced a FBA event was 15.3%, but the percentage of those who had intervened in such cases at least once was 6.9%. 28.4% of participants have knowledge and abilities in first aid interventions for FBA. Furthermore, 19.3% of participants possessed a first aid certification, although 59.6% indicated a readiness to undergo first aid training. The participants' average knowledge score was 11.68±4.32, signifying a moderate degree of knowledge level (Table 1). The distribution of participants across knowledge-level categories was as follows: 17.04% were classified as inadequate, 46.02% as moderate, and 36.94% as adequate (Figure 2).

The analysis of the correlation between participants' knowledge level scores and their socio-demographic features, as well as their knowledge and practices about first aid, reveals no statistically significant difference in the median scores among parents. Upon analysis by age groups, the median score for the 31-44 age cohort was elevated, with no statistically significant differences identified. The analysis revealed that the median score of secondary school graduates was elevated, while that of elementary school graduates was reduced, with no statistically significant difference identified. The analysis revealed that the median scores of waiters and managers were elevated and not statistically significant. No statistically significant difference exists between the median scores based on years of employment at the workplace. No statistically significant variation was seen in the median ratings based on the number of employees at the workplace. The median score of those possessing a first aid certificate is elevated, and there exists no statistically significant disparity. The median score of individuals inclined to obtain a first aid certificate is elevated in the response "I am undecided," and it lacks statistical significance. Significant differences were observed in knowledge level scores related to the possession of knowledge and skills in first aid practices ($p=0.001$, $r=0.26$), prior experience with a FBA incident ($p=0.036$, $r=0.16$), intervention in such incidents ($p=0.040$, $r=0.15$), and possession of knowledge and skills in first aid interventions for FBA ($p<0.001$, $r=0.34$) (Table 2).

The findings indicate that there is no statistically significant difference in the median attitude scores of participants when compared to their socio-demographic characteristics, knowledge, and practices related to first aid. The analysis indicated that the median score for high school graduates was elevated, while the median score for literate individuals was lower, with no statistically significant difference observed. No statistically significant difference was observed in the median scores based on the number of employees in the workplace. The median score of individuals who intervened in FBA incidents was higher, though this difference was not statistically significant. The median score of individuals willing to obtain a first aid certificate was higher for the statement "I don't want to," although this difference was not statistically significant. Significant differences in attitude scores were observed based on age group ($p=0.012$, $\eta^2=0.04$), job position ($p=0.022$, $\eta^2=0.03$), and years of employment ($p=0.018$, $\eta^2=0.03$). No statistically significant difference was observed when comparing the attitude scores of participants regarding FBA incidents and their willingness to receive first aid training. Significant differences in attitude scores were observed based on knowledge and skills in first aid practices ($p<0.001$, $\eta^2=0.07$), intervention during a FBA incident ($p=0.002$, $\eta^2=0.05$), knowledge and skills in first aid interventions for FBA ($p<0.001$, $\eta^2=0.08$), and possession of a first aid certification ($p=0.040$, $\eta^2=0.02$) (Table 3).

Participants exhibited a low percentage of correct responses to specific key statements in the knowledge assessment survey. A notable proportion of respondents failed to correctly identify the key signs of complete airway obstruction, including inability to breathe, clutching the neck, inability to speak, and cyanosis. Among the items with the lowest correct response rates, only 2.3% of participants correctly identified the proper positioning of an infant during a choking intervention.

Table 1. Distribution of data about the socio-demographic characteristics of participants and their knowledge and habits related to first aid		
Age, years (median, minimum-maximum)		34 (17-70)
Years of employment (median, minimum-maximum)		8 (0-40)
Number of employees in the workplace (median, minimum-maximum)		25 (1-343)
Knowledge level score (mean ± SD)		11.68±4.32
Attitude score (mean ± SD)		50.06±7.26
		n (%)
Gender	Male	139 (79.0)
	Female	37 (21.0)
Age group	17-30 years	74 (42.0)
	31-44 years	72 (41.0)
	45-70 years	30 (17.0)
Educational status	Literate	6 (3.4)
	Primary school	16 (9.1)
	Secondary school	30 (17.0)
	High school	98 (55.7)
	University	26 (14.8)
Job position	Waiter	70 (39.8)
	Kitchen staff	30 (17.0)
	Busboy	14 (8.0)
	Business owner	12 (6.8)
	Cashier	10 (5.7)
	Other	19 (10.8)
	Manager/supervisor	21 (11.9)
Years of employment	0-8 years	89 (50.6)
	9-20 years	58 (33.0)
	21-40 years	29 (16.5)
Number of employees in the workplace	1-14 employees	65 (36.9)
	15-40 employees	51 (29.0)
	41-343 employees	60 (34.1)
Possession of knowledge and skills in first aid practices	No	36 (20.5)
	Partially	84 (47.7)
	Yes	56 (31.8)
Encountering a foreign body aspiration incident	No	149 (84.7)
	Yes	27 (15.3)
Number of interventions for foreign body aspiration incidents	Never intervened	164 (93.1)
	Once	7 (4.0)
	Twice	4 (2.3)
	Three times	1 (0.6)
Possession of first aid knowledge and skills related to foreign body aspiration incidents	No	41 (23.3)
	Partially	85 (48.3)
	Yes	50 (28.4)
Possession of a first aid certification	No	142 (80.7)
	Yes	34 (19.3)
Willingness to receive first aid training	Do not want	33 (18.8)
	Undecided	38 (21.6)
	Want	105 (59.6)

SD: Standard deviation

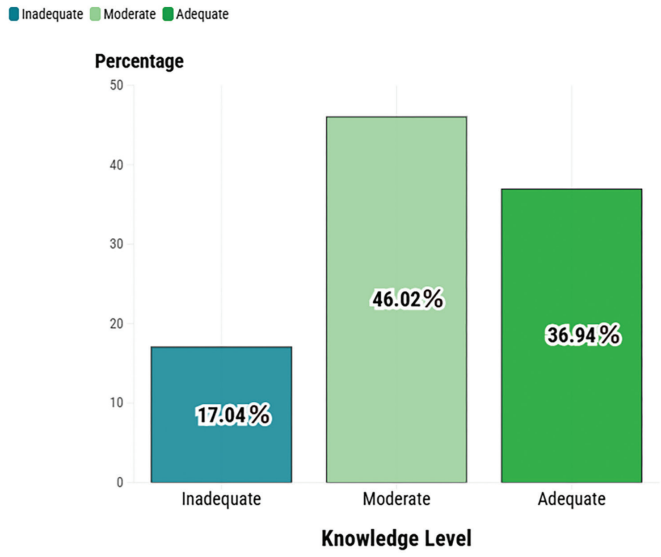


Figure 2. Distribution of participants by knowledge level scores

Table 2. Analysis of participants' knowledge level scores in relation to their socio-demographic factors and conclusions of their knowledge and behaviors regarding first aid

		Median (Q_1-Q_3)	p
Gender	Male	12 (9-15)	0.364
	Female	12 (7-14.5)	
Age group	17-30 years	12 (8-15)	0.067
	31-44 years	13 (11-15)	
	45-70 years	11 (8.5-14)	
Educational status	Literate	11 (4.75-17.5)	0.107
	Primary school	9 (7-13.25)	
	Secondary school	13 (8.75-14.25)	
	High school	12.5 (10.75-15)	
	University	11.5 (8.75-15)	
Job position	Waiter	13 (10-16)	0.159
	Kitchen staff	12 (9-15)	
	Busboy	9 (4.25-13)	
	Business owner	11.5 (7.5-14)	
	Cashier	11.5 (6.75-15.25)	
	Other	11 (7-14)	
	Manager/supervisor	13 (10-15)	
Years of employment	0-8 years	12 (8-15)	0.637
	9-20 years	12 (9-15)	
	21-40 years	12 (11-15)	
Number of employees in the workplace	1-14 employees	12 (8.5-15.5)	0.696
	15-40 employees	12 (9-15)	
	41-343 employees	11.5 (10-14)	
Possession of knowledge and skills in first aid practices	No[1]	10.5 (6-13.75)	**0.001****
	Partially[2]	12 (9-14.75)	
	Yes[3]	13 (11-16)	1<3

Table 2. Continued

		Median (Q_1-Q_3)	p
Encountering a foreign body aspiration incident	No	12 (9-15)	**0.036***
	Yes	14 (11-16)	
Number of interventions in foreign body aspiration incidents	Never intervened	12 (9-15)	**0.040***
	Intervened (includes 1, 2 and 3 times)	14.5 (12.25-15.75)	
Possession of first aid knowledge and skills in foreign body aspiration incidents	No[1]	10 (6-13)	**<0.001*****
	Partially[2]	12 (9-15)	1<2
	Yes[3]	14.5 (11.75-16)	1<3 2<3
Possession of a first aid certification	No	12 (9-15)	0.147
	Yes	13 (11-15)	
Willingness to receive first aid training	Do not want	13 (9-15)	0.201
	Undecided	13.5 (9.5-16)	
	Want	12 (9-14)	

Q_1: 25th percentile, Q_3: 75th percentile, *p<0.05, **p<0.01, ***p<0.001
Superscripts ([1]–[7]): Subgroup numbering for post-hoc comparisons. "1<2" indicates a significantly lower median in subgroup 1 than in subgroup 2

Table 3. Analysis of participants' attitude ratings in relation to their socio-demographic factors and findings of their knowledge and behaviors on first aid

		Median (Q_1-Q_3)	p
Gender	Male	50 (47-55)	0.607
	Female	51 (46.5-56)	
Age group	17-30 years[1]	49 (44.75-53)	**0.012***
	31-44 years[2]	52 (48-56)	
	45-70 years[3]	52 (47-57.25)	1<2
Educational status	Literate	48 (32-51.25)	0.479
	Primary school	50.5 (39.75-54)	
	Secondary school	50.5 (44.75-57)	
	High school	51 (47-55.25)	
	University	50 (48-54.5)	
Job position	Waiter[1]	51 (47-56)	**0.022***
	Kitchen staff[2]	51 (49-56)	3<1
	Busboy[3]	46.5 (39-50.75)	3<7
	Business owner[4]	53.5 (49.25-58.5)	3<2
	Cashier[5]	50 (44-54)	3<4
	Other[6]	49 (42-52)	6<2
	Manager/supervisor[7]	52 (48-56.5)	6<4
Years of employment	0-8 years[1]	47 (40.25-50.75)	**0.018***
	9-20 years[2]	50 (45.5-53)	
	21-40 years[3]	46 (38-54)	1<3
Number of employees in the workplace	1-14 employees	51 (48-56)	0.487
	15-40 employees	50 (46-55)	
	41-343 employees	50.5 (47-55)	

Table 3. Continued

		Median (Q_1-Q_3)	p
Possession of knowledge and skills in first aid practices	No[1]	47 (40.25-51.75)	**<0.001*****
	Partially[2]	51 (48-54)	1<2
	Yes[3]	52 (48-57)	1<3
Encountering a foreign body aspiration incident	No	50 (47-55)	0.316
	Yes	52 (45-58)	
Number of interventions in foreign body aspiration incidents	Never intervened	57 (46-60)	0.002**
	Intervened (includes 1, 2 and 3 times)	62 (55.75-64.5)	
Possession of first aid knowledge and skills in foreign body aspiration incidents	No[1]	47 (42-51.5)	**<0.001*****
	Partially[2]	51 (47.5-55)	1<2
	Yes[3]	52 (49-58.25)	1<3
Possession of a first aid certification	No	50 (46-55)	**0.040***
	Yes	52 (49-59)	
Willingness to receive first aid training	Do not want	51 (47-57)	0.405
	Undecided	50.5 (47-56)	
	Want	50 (47-54)	

Q_1: 25th percentile, Q_3: 75th percentile, *p<0.05, **p<0.01, ***p<0.001
Superscripts ([1]–[7]): Subgroup numbering for post-hoc comparisons. "1<2" indicates a significantly lower median in subgroup 1 than in subgroup 2

Additionally, just 4.0% recognized that FBA can cause fatal airway obstruction in all age groups. Furthermore, only 6.2% of participants accurately recognized that the abdominal thrust maneuver necessitates firm backward and upward pressure exerted with the hands. Additionally, prevalent misconceptions were noted concerning the age group most commonly affected, the significance of children maintaining concentration during meals, and the risks associated with walking or running while eating. Significant misconceptions were identified in first aid practices, particularly regarding the appropriate order of back blows and chest compressions. The findings underscore the necessity for specialized first aid training, with an emphasis on identifying symptoms of airway obstruction and executing suitable first aid techniques (Table 4).

The table below shows the response rates for the attitude survey, classified into the categories of "Strongly agree," "Agree," "Neutral," "Disagree," and "Strongly disagree" (Table 5).

Discussion

This study provides essential insights into an often neglected aspect of public health preparedness: the ability of frontline restaurant personnel to respond effectively to life-threatening choking incidents. Only 6.9% of participants reported having intervened in high-risk food service emergencies, and fewer than one-third indicated possessing the requisite first aid skills. The low prevalence of first aid certification, combined with a significant willingness to undergo training (59.6%), highlights the unmet demand and untapped potential for community-based interventions. These findings highlight the potential value of integrating context-specific, scenario-based first aid training into workplace safety programs, particularly in settings where FBA may occur.

Participants with both knowledge and practical skills in first aid exhibited significantly higher knowledge scores than those lacking such training. Furthermore, individuals possessing full or partial knowledge and skills exhibited more favorable attitudes towards first aid practices compared to those without any training. The findings indicate that theoretical understanding and practical experience are significant factors in emergency preparedness. Previous studies have reported similar associations, indicating that first aid training is correlated with enhanced knowledge and attitudes regarding emergency response (9,10). In the current study, only 28.4% of respondents reported possessing adequate first aid knowledge and skills, highlighting a significant deficiency in public preparedness. The disparity was notably pronounced among individuals with lower educational attainment, underscoring the necessity for enhanced access to first aid education, particularly via community-based initiatives.

Experiencing a real-life choking episode correlated with elevated knowledge scores among participants, although this experience seemingly did not affect their attitudes towards emergency response. This conclusion aligns with other studies indicating that direct exposure to emergencies may improve information retention but does not necessarily influence confidence or the propensity to act (11,12).

Table 4. Responses of participants to the knowledge assessment survey

	Incorrect	Don't know	Correct
	n (%)	n (%)	n (%)
Coughing and grasping the neck may indicate foreign body aspiration into the airways.	16 (9.1)	23 (13.1)	137 (77.8)
In cases of partial airway obstruction, where the individual is attempting to remove the foreign object by coughing, back blows should be administered to facilitate its clearance.	30 (17.1)	21 (11.9)	125 (71.0)
If the individual is attempting to remove the foreign object by coughing, they should be urged to persist in coughing.	34 (19.3)	36 (20.5)	106 (60.2)
Dyspnea, neck gripping in distress, aphonia, and cyanosis are indicators of total airway obstruction.	148 (84.1)	21 (11.9)	7 (4.0)
Intrusion of foreign items, including food particles, plastic, and metal, into the airways can result in life-threatening blockage across all age groups.	153 (86.9)	16 (9.1)	7 (4.0)
Foreign body aspiration in the airway is predominantly reported in adults across all age demographics.	77 (43.8)	37 (21.0)	62 (35.2)
Children should not be interrupted during meals to ensure adequate chewing, swallowing, and focus.	147 (83.6)	21 (11.9)	8 (4.5)
There is no harm in walking or running while eating.	140 (79.6)	18 (10.2)	18 (10.2)
Interruption of oxygen flow to the brain for over five minutes during foreign body aspiration episodes can result in brain damage and mortality.	149 (84.6)	20 (11.4)	7 (4.0)
Rescue interventions for infants and young children are the same as for adults.	25 (14.2)	35 (19.9)	116 (65.9)
The initial step should be to have the individual consume water.	42 (23.9)	25 (14.2)	109 (61.9)
The head should be tilted backward.	84 (47.7)	38 (21.6)	54 (30.7)
A maximum of three back blows should be administered in a sweeping motion.	28 (15.9)	35 (19.9)	113 (64.2)
To initiate the intervention, the rescuer should position themselves behind the individual and secure their torso.	6 (3.4)	20 (11.4)	150 (85.2)
One hand should be clenched into a fist and positioned just below the sternum, while the other hand provides support.	6 (3.4)	31 (17.6)	139 (79.0)
The hands must be forced forcefully backward and upward, with the thumb exerting pressure on the abdomen.	133 (75.6)	32 (18.2)	11 (6.2)
The abdominal thrust procedure must be performed repeatedly until the item is dislodged or the individual becomes unconscious.	21 (12.0)	52 (29.5)	103 (58.5)
When interfering with a newborn, the baby should be positioned face down on the arm, with the head supported and inclined forward.	134 (76.1)	38 (21.6)	4 (2.3)
A maximum of three rapid and vigorous back punches should be delivered between the infant's shoulder blades in a sweeping manner.	30 (17.0)	38 (21.6)	108 (61.4)
Following the administration of back blows, it is imperative to ascertain whether the airway is unobstructed.	7 (4.0)	30 (17.0)	139 (79.0)
If the item does not dislodge following back strikes, the child should be positioned supine with head support.	42 (23.9)	68 (38.6)	66 (37.5)
Position two fingers directly behind the sternum and provide five compressions.	14 (7.9)	70 (39.8)	92 (52.3)
The compression maneuver must be sustained until the foreign object is evacuated or the child becomes unconscious.	24 (13.7)	71 (40.3)	81 (46.0)
An individual may execute the move independently.	49 (27.9)	37 (21.0)	90 (51.1)
They should form a fist and position it behind the sternum.	29 (16.4)	55 (31.3)	92 (52.3)
They should apply pressure to the abdomen by pressing their fist against a solid surface, such as the back of a chair.	14 (8.0)	50 (28.4)	112 (63.6)

Table 5. Responses of participants to the attitude survey

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
	n (%)	n (%)	n (%)	n (%)	n (%)
I anticipate that I could witness an individual choking as a result of foreign body aspiration in everyday situations.	67 (38.1)	59 (33.5)	34 (19.3)	5 (2.8)	11 (6.3)
It is my belief that all individuals should possess the capability to execute first aid interventions.	83 (47.2)	41 (23.3)	15 (8.5)	18 (10.2)	19 (10.8)
I argue that restaurant personnel ought to undergo first aid training.	99 (56.3)	51 (29.0)	11 (6.3)	4 (2.3)	11 (6.3)
Consuming meals with bones or fishbones without meticulous removal poses no risk.	96 (54.5)	54 (30.7)	26 (14.8)	0 (0.0)	0 (0.0)
I am certain that I possess the information to intervene.	22 (12.5)	40 (22.7)	76 (43.2)	20 (11.4)	18 (10.2)
I am capable of intervening in such a case.	25 (14.2)	50 (28.4)	69 (39.2)	17 (9.7)	15 (8.5)
I might be unable to intervene out of panic.	42 (23.9)	69 (39.2)	65 (36.9)	0 (0.0)	0 (0.0)
I may assess if an individual's airway is partially or fully clogged by observation.	21 (11.9)	42 (23.9)	84 (47.7)	15 (8.5)	14 (8.0)
I believe that I should call 112 immediately.	67 (38.1)	53 (30.1)	20 (11.4)	22 (12.5)	14 (8.0)
I should not intervene until 112 arrives.	78 (44.3)	59 (33.5)	39 (22.2)	0 (0.0)	0 (0.0)
I believe that the person should be assisted immediately.	76 (43.2)	54 (30.7)	26 (14.8)	8 (4.5)	12 (6.8)
I believe that the abdominal thrust maneuver is life-saving.	77 (43.8)	50 (28.4)	28 (15.9)	5 (2.8)	16 (9.1)
I think that striking the person's chest may be beneficial.	68 (38.6)	56 (31.8)	52 (29.5)	0 (0.0)	0 (0.0)

Participants who intervened in FBA incidents demonstrated significantly greater knowledge and attitude scores than those who did not intervene. This finding indicates that participation in real-life emergencies may improve cognitive comprehension and attitudinal preparedness for first aid. The findings align with prior research, highlighting the reinforcing impact of direct involvement on emergency preparedness (13). Only 6.9% of participants in our study reported intervening in such situations, reflecting a low level of active bystander participation.

The scores for knowledge and attitude were positively correlated with the level of proficiency in first aid. Participants possessing comprehensive first aid knowledge and skills achieved the highest scores, followed by those with partial knowledge, and finally those lacking any training. This sequential pattern reinforces the hypothesis that perceived self-efficacy in first aid is associated with increased knowledge and more positive attitudes. Our findings align with existing literature that similarly associates self-reported competence with enhanced preparedness for emergency situations (14,15). Only 31.8% of participants reported having first aid knowledge and skills. This suggests that both training and opportunities for practical reinforcement —such as scenario-based exercises— may help strengthen preparedness.

No statistically significant differences in knowledge or attitude scores were found between genders, consistent with previous research indicating that gender does not substantially influence first aid knowledge or attitudes (16). Approximately 80% of participants were male, which may have limited the ability to detect potential gender-related differences. Future studies with more balanced samples could help clarify whether gender plays a meaningful role in first aid preparedness.

No statistically significant differences were observed in knowledge scores across age groups. In contrast, attitude scores were significantly higher among participants aged 31-44 compared with those aged 17-30. This pattern is consistent with previous literature indicating that age does not necessarily enhance factual first aid knowledge, yet may be associated with more positive attitudes toward emergency response behaviors (17,18). The age distribution in our sample was relatively balanced—42.0% aged 17-30 and 41.0% aged 31-44—suggesting that the observed difference is unlikely to be attributable to unequal group sizes. The more favorable attitudes in the older cohort may reflect accumulated life experience, including greater exposure to emergencies or caregiving roles, which could enhance perceived responsibility or willingness to intervene in FBA events.

No significant differences in either knowledge or attitude scores were identified across educational levels. This finding is aligned with previous studies reporting that formal educational attainment alone does not reliably predict first aid competence or readiness to intervene (12,19). In our sample, more than half of participants (55.7%) had completed only secondary education, and no significant differences in knowledge or attitude scores were identified across educational levels. These findings suggest that first aid readiness may be independent of formal educational attainment. Accordingly, first aid training programs should be designed to remain accessible and applicable to individuals with varied educational backgrounds.

No significant differences in knowledge levels were seen among occupational occupations; however, attitude ratings exhibited considerable variation. Waiters, managers/directors, culinary workers, and business owners exhibited markedly superior attitude scores compared to bellboys. Furthermore, kitchen personnel and proprietors had elevated attitude ratings relative to other occupational groups. The findings correspond with previous studies indicating that employment status may influence individuals' attitudes about first aid, possibly owing to differences in duties, perceived accountability, and direct engagement in high-risk situations (15). Considering that roughly fifty percent of our sample comprised waiters, this subgroup constitutes a pivotal target for intervention. The variation in attitude scores across occupational groups suggests that training needs may differ among staff roles. Therefore, first aid programs in food service settings should be implemented in a way that ensures all personnel receive consistent and role-appropriate instruction.

No significant differences in knowledge scores were seen according to years of work. Participants with 21-40 years of work experience had markedly superior attitude ratings in comparison to those with 0-8 years of experience. These findings align with other research suggesting that professional tenure may not directly impact theoretical understanding, although might positively alter attitudes toward emergency response (20). Differences in attitude scores across experience groups suggest that longer employment may be associated with greater comfort or willingness to intervene, although this study was not designed to determine causality. The absence of variation in knowledge scores across experience levels indicates that work tenure alone may not influence factual first aid knowledge.

Participants exhibited a moderate level of first aid knowledge. This aligns with findings from other sectors. Previous studies have indicated intermediate levels of first aid knowledge among cab drivers (21). Although many studies indicate that persons with previous first aid training often have superior knowledge ratings, they may nevertheless see themselves as unprepared or inefficient in real emergency situations (22). In a separate research, 27.1% of participants characterized their first aid knowledge as moderate (23), reinforcing the idea that both the quality of training and knowledge retention are essential for effective emergency response.

Although holding of a first aid certification did not correlate with significantly elevated knowledge scores, certified individuals had more favorable opinions regarding first aid procedures. In our survey, around 20% of participants possessed a valid first aid certification. This implies that although certification by itself may not ensure superior information retention, it may enhance confidence and preparedness to take action. The limited overall perceptions identified in the study highlight the need to improve the accessibility and structure of first aid education. No significant differences in knowledge or attitude ratings were detected about participants' readiness to undergo first aid instruction. Nonetheless, over half (59.6%) indicated a desire to participate in such training. This corresponds with previous research, wherein participants indicated unhappiness with earlier first aid training and stated a want for re-education (24). These observations present a significant potential for public health initiatives to leverage community motivation by enhancing the accessibility and attractiveness of first aid programs. This interpretation is consistent with international evidence: a quasi-experimental study from Indonesia demonstrated that community-based choking first aid education resulted in substantial improvements in laypersons' knowledge, skills, and intervention confidence, underscoring the transformative effect of structured, scenario-based training models in diverse populations (25).

This study enhances the Turkish public health literature by concentrating on the preparation of restaurant staff for FBA events, a subject usually examined in pediatric or clinical contexts. This research examines a high-risk, frequently neglected occupational category that routinely engages with the public in environments where choking incidents are probable, contrasting with other studies that focused on legally obliged professions such as educators or healthcare workers. This study contributes to the limited evidence on the preparedness of non-medical frontline workers to respond to choking incidents by examining their first aid knowledge and attitudes. To our knowledge, it is the first district-based study in Türkiye focusing on restaurant staff in this context. The findings provide baseline data that may inform future workplace training strategies and community-oriented public health initiatives.

Study Limitations

This study has several limitations. First, it was conducted in a single district of Istanbul, which restricts the external validity of the findings and limits their applicability to other regions or restaurant settings. Second, data collection occurred during a single visit to each establishment, and only employees who were present at that moment were included. This may have introduced selection bias, as individuals with different shifts, levels of experience, or varying preparedness may not have been represented. Third, all responses were based on self-report, which is susceptible to social desirability and recall bias, potentially leading participants to overestimate their knowledge or confidence. Additionally, the knowledge assessment instrument—although developed using national guidelines and expert opinion—has not undergone formal psychometric validation, as no standardized tool exists for measuring FBA first aid knowledge. Finally, the study did not incorporate direct observation or practical skill assessment; therefore, the results reflect perceived rather than demonstrated first aid competence.

Conclusion

This study underscores a significant deficiency in first aid readiness among restaurant personnel in addressing FBAO occurrences. Despite operating in high-risk settings where prompt intervention is critical, only a little percentage of participants possessed genuine first aid certification, and an even smaller fraction had previously intervened in choking incidents. Despite over half stating a readiness to undergo training, this has not resulted in sufficient preparedness, highlighting a disparity between intention and actual readiness.

Considering that restaurant staff frequently serve as the initial and only responders in crises, it is essential to incorporate mandatory, scenario-based first aid and basic life support training into occupational safety regulations. Broadening access to this training outside the food service industry—via community-oriented public health initiatives—could also elevate bystander intervention rates and increase survival outcomes in FBAO incidents.

First aid education must be rendered more accessible, consistently updated, and tailored to the requirements of non-healthcare professionals. As rules and norms develop, continuous refresher training for all occupational groups is essential to sustain competency and avert skill deterioration over time.

These findings highlight the necessity of including systematic, ongoing, and pragmatic first aid training within workplace safety rules and comprehensive health education initiatives. Legal certification requirements, bolstered by simulation-based learning, may improve information retention and the confidence necessary for successful action in real-life situations.

This study effectively fills a significant gap in the literature by examining the preparation of restaurant personnel in FBAO circumstances. Future study should investigate obstacles to training access—such as cost, time, and availability—and assess the long-term effects of tailored interventions in both professional and community contexts. Addressing these concerns can enhance the capacity of laypersons as effective first responders, hence improving emergency outcomes and fostering community resilience.

Ethics

Ethics Committee Approval: Prior to the commencement of the study, written approval was obtained from the University of Health Sciences Türkiye Hamidiye Scientific Research Ethics Committee (decision no.: 2023/3, date: 10.02.2023). Throughout the study, the principles of the Declaration of Helsinki were strictly adhered to.

Informed Consent: Informed consent was obtained from all participants prior to their inclusion in the study.

*This study was presented as an oral presentation at the Third National Congress on Holistic Approach to Medicine on October 7-8, 2020 (pages: 268-272).

Footnotes

Authorship Contributions

Concept: M.T.U., F.Y., G.A., H.K., M.N.Ü., G.B., Design: M.T.U., F.Y., G.A., H.K., M.N.Ü., G.B., Data Collection or Processing: M.T.U., F.Y., E.Ç., G.A., H.K., M.N.Ü., Analysis or Interpretation: M.T.U., E.Ç., Literature Search: M.T.U., F.Y., E.Ç., G.A., H.K., M.N.Ü., G.B., Writing: M.T.U., F.Y., E.Ç., G.A., H.K., M.N.Ü., G.B.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Güzel A, Paksu S. Çocuk ve yetişkinlerdeki yabancı cisim aspirasyonları. J Exp Clin Med. 2013;30(Suppl 1):81-5.
2. Ebrahimi M, Mirhaghi A. Heimlich maneuver complications: a systematic review. Eurasian J Emerg Med. 2019;18:157-65.
3. Berzlanovich AM, Fazeny-Dörner B, Waldhoer T, Fasching P, Keil W. Foreign body asphyxia: a preventable cause of death in the elderly. Am J Prev Med. 2005;28:65-9.
4. Soroudi A, Shipp HE, Stepanski BM, Ray LU, Murrin PA, Chan TC, Adult foreign body airway obstruction in the prehospital setting. Prehosp Emerg Care. 2007;11:25-9.
5. Ucar MT, Dönmez E. Epidemiological insights from media reports on foreign body aspiration in the trachea: a retrospective descriptive study. Eurasian J Med Investig. 2025;9:79-88.
6. Ulas AB, Aydin Y, Eroglu A. Foreign body aspirations in children and adults. Am J Surg. 2022;224:1168-3. Epub 2022 May 27.
7. Kim IA, Shapiro N, Bhattacharyya N. The national cost burden of bronchial foreign body aspiration in children. Laryngoscope. 2015;125:1221-4.
8. İnan HF, Kurt Z, Kubilay İ. Basic first aid practices training book. Ankara: Republic of Turkey Ministry of Health; 2011. Available from: https://www.ilkyardim.org.tr/dokumanlar/Saglik-Bakanligi-Ilk-Yardim.pdf.
9. Karadeniz Coşar B. First aid knowledge level and awareness among adults in Efeler district of Aydın [Specialist thesis]. Aydın: Aydın Adnan Menderes University, Faculty of Medicine, Department of Family Medicine; 2023. Coşar B.
10. Semiz İ. Assessment of first aid knowledge among employees in the furniture manufacturing sector: the case of İnegöl [Master's thesis]. Istanbul: Istanbul Rumeli University, Institute of Graduate Education, Department of Occupational Health and Safety; 2023.
11. Özdemir E. Mothers' safety measures for home accidents and first aid self-efficacy [Master's thesis]. Çanakkale: Çanakkale Onsekiz Mart University, Graduate School of Education; 2023
12. Gülser Uruk M. First aid knowledge and survival preparedness of underground mine workers during disasters [Master's thesis]. Istanbul: Bezmialem Vakif University, Institute of Health Sciences, Department of Disaster Management; 2020.
13. Coşkun C. First aid knowledge levels and frequency of first aid situations among mothers with children aged 0–14 in Çankırı-Eldivan [Master's thesis]. Ankara: Gazi University, Institute of Health Sciences, Department of Demographics and Epidemiology of Accidents; 2004.
14. Ergen Gürarslan N. First aid knowledge levels and frequency of first aid conditions among mothers with children aged 0–14 in Tunceli-Pertek [Master's thesis]. Elazığ: Fırat University, Institute of Health Sciences, Department of Public Health; 2012.
15. Aytaç Ş. Frequency of first aid situations and changes in knowledge before and after first aid training among workers in a furniture factory in Ankara-Akyurt [Master's thesis]. Ankara: Gazi University, Institute of Health Sciences, Department of Occupational Health and Safety; 2010.
16. Perişan Eker K. The relationship between health literacy and basic first aid knowledge among driver's license applicants [Specialist thesis]. Istanbul: University of Health Sciences, Başakşehir Çam and Sakura City Hospital, Department of Family Medicine; 2024.
17. Kurt FY, Aytekin A. Home accidents among children aged 0–6 years. J Health Sci Professions. 2015;2:22-32.
18. Deniz R. Determining the qualifications of personnel to receive first aid training in industrial workplaces in Bursa [Master's thesis]. Istanbul: Okan University, Health Sciences Institute, Department of Health Management; 2014.
19. Karadağ T. Determining the level of knowledge about first aid of Turkey's construction sector employees. Kent Akademisi. 2023;16:2015-27.
20. Yağan A, Şahin HA. Evaluation of factors affecting first aid and basic life support knowledge in hazardous workplaces. Ahi Evran Med J. 2023;7:280-7.
21. Ardoğan O. The importance of first aid and emergency services and first aid knowledge, attitudes, and behaviors of taxi drivers in Ankara [Master's thesis]. Ankara: Gazi University, Institute of Science, Department of Environmental and Technical Research of Accidents; 1998
22. Toraman F. First aid knowledge of family caregivers responsible for children with cerebral palsy: the case of Yalova [Master's thesis]. Yalova: Yalova University, Graduate Education Institute, Department of Social Work; 2024.
23. Babaoğlu C. Evaluation of the effectiveness of face-to-face and online first aid training provided to workers at a higher education institution [Master's thesis]. Kutahya: Kutahya Health Sciences University, Graduate Education Institute, Department of Nursing; 2022.
24. Çağlayan H. First aid knowledge levels and attitudes of employees in medium-sized enterprises (SMEs) in Tokat [Master's thesis]. Tokat: Tokat Gaziosmanpaşa University, Graduate Education Institute, Department of Occupational Health and Safety; 2023.
25. Affrita TM, Kusuma F, Warasanti ES, Tanggono A. Effectiveness of a community-based first aid education program on choking for laypersons in Surabaya, Indonesia: a quasi-experimental study. Arkus. 2024;10:647-58.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.02222
Eurasian J Emerg Med. 2026;25: 111-6

Assessment of the Educational Value and Technical Competence of Surgical Incision Repair Videos Shared on YouTube

Fatih Gönültaş

İnönü University Faculty of Medicine, Department of General Surgery, Malatya, Türkiye

Abstract

Aim: This study aimed to evaluate the educational quality, reliability, and technical accuracy of surgical incision repair videos on YouTube and to investigate the influence of video source, country of origin, and content characteristics on overall quality scores.

Materials and Methods: All videos were independently assessed by two researchers using four validated scoring systems: the Journal of the American Medical Association (JAMA) criteria, modified DISCERN, Global Quality score (GQS), and objective structured assessment of technical skills (OSATS). Video popularity was calculated using the video power index. Statistical analyses were conducted using the Mann-Whitney U test, Spearman's correlation analysis, and ROC curve evaluation, with statistical significance set at $p<0.05$.

Results: According to the DISCERN classification, 48.5% of videos were of very poor quality, while 12.1% were categorized as good quality. Institutional videos achieved significantly higher scores across all evaluation systems compared with individual uploads ($p<0.05$). Videos with spoken narration received higher scores for educational and technical quality than silent videos. Content originating from the United States had significantly higher JAMA, GQS, and DISCERN scores than content from other countries ($p<0.05$). Positive correlations were identified between OSATS scores and JAMA ($r=0.315$), GQS ($r=0.782$), and DISCERN ($r=0.702$) scores.

Conclusion: The overall educational quality of YouTube videos on surgical incision repair was moderate. Institutional and narrated videos yielded significantly higher quality scores, underscoring the importance of academic oversight and structured content development. Although YouTube represents a valuable supplementary tool for surgical education, standardization and scientific regulation are essential to ensure educational reliability.

Keywords: Surgical incision repair, educational quality, JAMA, modified DISCERN, Global Quality score, OSATS, video power index

Introduction

Surgical incision repair is a fundamental skill that directly influences wound healing, infection risk, and aesthetic outcomes. Optimal repair requires adherence to basic surgical principles, including appropriate suture selection, balanced tissue tension, elimination of dead space, and minimization of tissue trauma (1,2). Consequently, acquiring proper incision repair technique constitutes an essential component of surgical education for medical students and residents.

The rapid digital transformation in medical education has increased the use of online video platforms as supplementary learning tools. YouTube, being freely accessible and visually rich, has become one of the most frequently used platforms for observing surgical techniques. However, the uncontrolled nature of user-generated content raises concerns regarding the accuracy, reliability, and educational adequacy of such videos (3-5).

To objectively evaluate online surgical content, several validated scoring systems have been introduced. The Journal of the American Medical Association (JAMA) criteria, the Global Quality score (GQS), and the modified DISCERN instrument assess the reliability and completeness of medical information (6,7). Furthermore, the objective structured assessment of technical skills (OSATS) has been utilized to evaluate the technical quality of surgical performance in video format (8). Prior studies have consistently demonstrated that videos produced by academic or institutional sources receive higher quality scores, whereas individually uploaded content

Corresponding Author: Fatih Gönültaş MD, İnönü University Faculty of Medicine, Department of General Surgery, Malatya, Türkiye
E-mail: fatnih44@gmail.com **ORCID ID:** orcid.org/0000-0001-7771-3891
Cite this article as: Gönültaş F. Assessment of the educational value and technical competence of surgical incision repair videos shared on YouTube. Eurasian J Emerg Med. 2026;25: 111-6.

Received: 08.10.2025
Accepted: 15.12.2025
Epub: 22.12.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

often lacks accuracy and completeness (9-11). These findings emphasize the need for standardized and structured evaluation of surgical educational videos.

The urgency of wound care underscores the critical role of both emergency departments and general surgery services in the timely management of acute injuries. Emergency physicians are often the first clinicians to assess traumatic or surgical wounds, and their initial decisions regarding incision, exploration, and early repair have direct consequences for infection control, functional recovery, and cosmetic outcomes. General surgeons subsequently provide definitive management and ensure that repair techniques adhere to established surgical principles.

The present study aims to assess the educational quality, reliability, and technical accuracy of videos of surgical incision repair available on YouTube. Additionally, the effects of video source (institutional vs. individual), country of origin, and presentation style on video quality scores were examined. The findings are intended to contribute to the growing body of literature on digital surgical education and highlight the need for scientifically supervised, high-quality online instructional content.

Materials and Methods

Video Selection

A comprehensive video search was conducted on the YouTube platform on September 10, 2025. To ensure methodological reproducibility, each keyword was searched separately, and the first 50 results for each keyword were recorded. In total, 250 videos (5 keywords × 50 results each) were initially screened. The keyword set-"surgical incision," "scalpel handling," "tissue dissection," "surgical cutting techniques," and "safe cutting in surgery" was deliberately selected to encompass the entire spectrum of surgical steps that directly precede or accompany incision repair. Although the study primarily focused on incision repair, videos demonstrating incision techniques were generally uploaded under broader surgical skill categories such as instrument use, tissue dissection, or general surgical cutting.

Duplicate and Irrelevant Video Elimination

All 250 initial videos were exported into a screening spreadsheet. Duplicates were identified by matching identical uniform resource locators, identical uploader names with repeated titles, and algorithmically overlapping videos across different keyword searches.

Irrelevant videos were excluded based on their title, description, and the first 60 seconds of content. Videos were removed if they did not include any live, cadaveric, or simulated incision or dissection step; consisted solely of animations without procedural depiction; focused exclusively on postoperative wound care or complications; or included advertisements, promotional content, or non-instructional material. After this multistep elimination, 66 videos met the predefined criteria and were included in the final analysis.

Inclusion and Exclusion Criteria

Inclusion Criteria Were:

1. Videos demonstrating step-by-step surgical incision or related dissection techniques, (justification added: Incision repair cannot be evaluated independently of the preceding incision technique; poor incision may influence subsequent repair quality).

2. Visual resolution ≥480p to ensure evaluable technical detail,

3. Real surgery, cadaveric, animal model, or validated simulation footage,

4. Availability of English narration or subtitles for accurate interpretation of procedural intent and terminology.

Exclusion Criteria Were

1. Promotional, humorous, or entertainment-oriented content,

2. Videos dedicated solely to postoperative care or complication management,

3. Videos with insufficient audio/visual quality preventing technical assessment,

4. Repetitive uploads from the same uploader to prevent content duplication bias.

Data Collection

For each video, the following parameters were recorded:

- Title, duration, upload date, number of views and likes,
- Video source: institutional (university, hospital, surgical society) or individual (surgeon, student, general user),
- Country of origin (United States, Türkiye, others),
- Type of presentation (spoken narration, silent, or subtitled).

The country of origin for each video was determined using a multi-step verification protocol. First, the uploader's profile information was examined, including self-declared institutional affiliation, professional designation, and geographic details. Channel metadata, such as language, institutional logos, and linked websites, was also reviewed. When this information was unclear, cross-verification was performed using publicly available data regarding the affiliated institution or surgeon featured in the video, official web pages linked to the YouTube channel, and

professional registry records, when applicable. If none of these sources provided reliable information, the video was classified as "others/not identifiable" to avoid misclassification bias.

All videos were independently assessed by two reviewers. In cases of scoring discrepancies, reviewers discussed the differences and reached agreement; because consensus-based final scores were used for statistical analyses, formal inter-rater reliability metrics could not be calculated. Each video was examined for the appropriateness of the surgical incision technique, including incision line planning, scalpel handling, tissue tension management, hemostasis, and tissue preservation.

Video Quality Assessment Systems

Each video was evaluated using four validated scoring systems:

1. JAMA Criteria:

Comprising four parameters-authorship, attribution, disclosure, and currency-each assigned one point, with a maximum possible score of 4.

2. Modified DISCERN:

A five-question instrument assessing information reliability and content integrity. Each "Yes" response was assigned 1 point, and each "no" response was assigned 0 points, yielding a maximum score of 5.

3. GQS:

A five-point scale evaluating overall flow, informativeness, and educational value, where 1 indicates poor and 5 indicates excellent quality.

4. OSATS:

Used to evaluate technical performance across five criteria:

- Scalpel handling,
- Tissue manipulation,
- Field of vision control,
- Hemostasis management,
- Safety and surgical planning.

Each item was rated from 1 to 5, with a maximum possible score of 25.

For technical evaluation, we used a video-adapted version of the OSATS. The selected OSATS domains—scalpel handling, tissue manipulation, field-of-vision control, hemostasis, and safety/surgical planning—represent components that can be objectively assessed from recorded videos rather than by real-time observation. Both reviewers were trained general surgeons with experience in surgical skills education. These surgeons independently assigned OSATS scores before reaching a consensus.

The mean total score obtained from these four systems was recorded as the overall video quality indicator.

Popularity Analysis

The duration, title, number of views, time elapsed since upload, viewing rate (views/day), number of comments, number of likes and number of dislikes, and like ratio [likes × 100 / (likes + dislikes)] were recorded for each video. Video popularity was assessed using the video power index (VPI), a standardized indicator of user engagement and visibility on the platform. The VPI was calculated using the formula: (like ratio × viewing rate) / 100, which incorporates the number of views, likes, and comments.

Statistical Analysis

All statistical analyses were performed using IBM SPSS Statistics version 27.0 (IBM Corp., Armonk, NY, USA). Continuous variables were expressed as median and interquartile range (IQR), while categorical variables were presented as percentages (%). Comparisons between two groups (e.g., institutional vs. individual videos) were performed using the Mann-Whitney U test. Correlations between video quality metrics were analyzed using the Spearman's correlation test. A p value <0.05 was considered statistically significant.

Results

Quantitative data for YouTube videos are presented in Table 1 as median and IQR. YouTube Global score [2069.5 (1509.2-2069.5)], subscribers [(5670 (196.7-70575)], likes [(222.5 (23.2-2500)], views [33001.1 (1618.7-239.682)], video duration [5.1 (2-8.7)], JAMA [1 (1-2)], GQS [3 (2.7-4)], DISCERN [2 (1-3)], OSAST [3 (2.7-3)], viewing rate [6.9 (1.2-48.2)], like ratio [0.07 (0.01-0.4)].

According to the DISCERN classification, 48.5% (32) of the videos were of very poor quality, 21.2% (14) were of poor quality, 18.2% (12) were of medium quality, and 12.1% (8) were of good quality. No videos of excellent quality met the validation criteria for this classification. According to the OSATS classification, 10.6% (7) of the videos were rated as low technical quality, 13.6% (9) as medium quality, 53% (35) as good quality, and 22.7% (15) as excellent quality.

When videos were categorised as corporate or individual, the numbers of subscribers and likes, and the JAMA, GQS, DISCERN, and Osast scores were statistically significantly higher for corporate videos. ($p=0.006$, $p=0.007$, $p=0.016$, $p=0.002$, $p=0.004$, $p=0.001$)

The videos were divided into two groups based on whether they contained spoken narration. The group containing spoken narration had statistically significantly higher values for subscriber count, like count, view count, video duration, and JAMA, GQS, DISCERN, and OSATS scores compared to the other group. (p=0.012, p=0.005, p=0.019, p=0.004, p=0.014, p=0.001, p=0.001, p=0.001)

The JAMA, GQS, and DISCERN scores for trending videos in the USA were significantly higher (p=0.032, p=0.001, and p=0.02).

When correlation analysis was performed using the OSATS Monitoring index and the JAMA, GQS, and DISCERN classifications, all indices showed a positive correlation with OSATS. The correlation analyses are presented in Table 2. When ROC analysis was performed with corporate and individual loaders using the JAMA, GQS, DISCERN, and OSATS scoring indices, the highest sensitivity (84.1%) was observed in corporate loaders with OSATS, followed by DISCERN (72.7%), GQS (61.4%), and JAMA (47.7). (Table 3, Figure 1)

Table 1. Quantitative data for YouTube videos: median and interquartile range

	Median	IQR	
		25%	75%
NDRA	2069.5	1509.2	2069.5
Subscriber	5670	196.7	70575
Likes	222.5	23.2	2500
View	33001.1	1618.7	239.682
Video duration	5.1	2	8.7
JAMA	1	1	2
GQS	3	2.7	4
DISCERN	2	1	3
OSAST	3	2.7	3
Viewing rate	6.9	1.2	48.2
Like ratio	0.07	0.01	0.4

NDRA: Number of days it remained on air, JAMA: Journal of American Medical Association, GQS: Global Quality score, OSAST: Objective structured assessment of technical skills

Table 2. Correlation analysis between the OSATS Monitoring index and the JAMA, GQS, DISCERN classifications

		r	p value
JAMA	OSATS	0.315*	0.01
GQS		0.782**	0.001
DISCERN		0.702**	0.001

*: Low, **: High, JAMA: Journal of American Medical Association, GQS: Global Quality score, OSAST: Objective structured assessment of technical skills

Discussion

In this study, we evaluated the educational quality, reliability, and technical accuracy of surgical incision repair videos on the YouTube platform. Overall, the moderate JAMA, modified DISCERN, GQS, and OSATS scores indicate that, although many videos provide basic instructional content, they often lack the depth, structure, and standardization expected in formal surgical education.

Consistent with prior research, institutional videos demonstrated significantly higher quality scores than individual uploads. Similar findings have been reported in studies assessing sleeve gastrectomy, laparoscopic prostatectomy, and resuscitation-related videos, where academic or professionally produced content consistently received higher scores on established quality indicators (7,9,12). These results underscore the continued influence of content source reliability on educational accuracy and completeness.

Our findings also showed that videos originating in the United States had higher overall quality and viewership metrics than those from other countries, which aligns with previous literature suggesting that North American and European institutions tend to produce more structured, professionally edited surgical videos (13). Nevertheless, the influence of geographic origin should be interpreted cautiously, as cross-country comparisons remain relatively limited in the current literature.

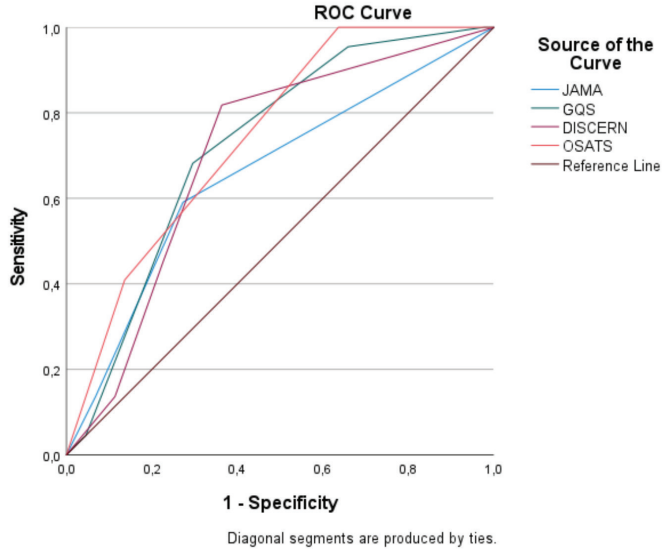


Figure 1. ROC graph according to JAMA, GQS, DISCERN, and OSATS scoring indices with corporate and individual loaders

JAMA: Journal of American Medical Association, GQS: Global Quality score, OSAST: Objective structured assessment of technical skills

Table 3. ROC analysis according to JAMA, GQS, DISCERN, and OSATS scoring indices for institutional and individual uploaders						
	AUC	p value	Cutoff	Sensitivite	Spesifite	Youden index
JAMA	0.66	0.035	1.5	47.7	81.8	0.295
GQS	0.803	0.001	3.5	61.4	95.5	0.568
DISCERN	0.839	0.001	1.5	72.7	90.9	0.636
OSATS	0.735	0.002	2.5	84.1	40.9	0.25
JAMA: Journal of American Medical Association, GQS: Global Quality score, OSAST: Objective structured assessment of technical skills, AUC: Area under the curve						

The moderate OSATS scores further highlight gaps in the technical accuracy of many videos. Suboptimal demonstrations of incision execution, scalpel handling, tissue manipulation, and hemostasis management may mislead inexperienced trainees and reinforce improper techniques. Although the LAP-VEGaS guidelines provide a structured framework for high-quality surgical video reporting (8), adherence appears limited among publicly available videos. Similar observations have been made in studies examining suturing, knot-tying, and instrument-handling videos, in which a substantial proportion of the content was found to be technically inadequate (12,14).

Another noteworthy finding was the lack of a strong correlation between popularity metrics, such as the VPI, and educational or technical quality. Videos with lower quality scores often received high view counts, reflecting a discrepancy between viewer engagement and scientific accuracy. Previous studies have similarly reported that popularity on YouTube does not necessarily reflect educational value (4,15).

Overall, the results of this study support the growing consensus that YouTube can serve as a complementary tool for surgical training; however, its lack of peer review, standardization, and quality control continues to limit its reliability as a stand-alone educational resource. Further efforts by academic institutions to develop and disseminate high-quality, peer-reviewed educational videos are essential for improving the pedagogical value of online surgical content (16,17).

Study Limitations

This study has several limitations. First, although searches were performed in incognito mode to minimize bias, YouTube's search algorithm is dynamic and personalized, and variations in displayed content cannot be fully eliminated. Second, restricting the analysis to English-language videos may have excluded potentially valuable content in other languages and may have limited the generalizability of the findings. Third, although a video-adapted OSATS approach was used, OSATS was originally developed for real-time observation, and certain nuances of technical performance may not be fully captured in recorded footage. Fourth, despite independent evaluations by two reviewers, subjective bias cannot be entirely excluded; because consensus scoring was applied, formal inter-rater reliability calculations could not be performed. Fifth, YouTube metadata such as likes, views, and comments are user-generated, non-standardized, and susceptible to external influences, which may limit the interpretability of these popularity metrics.

Conclusion

Finally, the rapidly evolving nature of YouTube content means that video availability and engagement statistics may change over time.

Ethics

Ethics Committee Approval: This study included only YouTube videos and did not involve any patient data or interventional procedures; therefore, ethical committee approval was not required.

Informed Consent: The study did not involve direct patient participation or the use of identifiable patient data; therefore, informed consent was not obtained.

Footnotes

Financial Disclosure: The author declared that this study received no financial support.

References

1. Baranoski, Sharon, and Elizabeth A. Ayello. Wound care essentials: practice principles. Lippincott Williams & Wilkins, 2008.
2. Singer AJ, Dagum AB. Current management of acute cutaneous wounds. N Engl J Med. 2008;359:1037-46.
3. Azer SA. Can "YouTube" help students in learning surface anatomy? Surg Radiol Anat. 2012;34:465-8.
4. Rapp AK, Healy MG, Charlton ME, Keith JN, Rosenbaum ME, Kapadia MR. YouTube is the most frequently used educational video source for surgical preparation. J Surg Educ. 2016;73:1072-6.
5. Maziriri ET, Gapa P, Chuchu T. Student perceptions towards the use of YouTube as an educational tool for learning and tutorials. 2020;13:119-38.
6. Rössler B, Lahner D, Schebesta K, Chiari A, Plöchl W. Medical information on the internet: quality assessment of lumbar puncture and neuroaxial block techniques on YouTube. Clinical Neurology and Neurosurgery. 2012; 114:655-8.

7. Arslan B, Gönültaş S, Gökmen E, Özman O, Onuk Ö, Yazıcı G, et al. Does YouTube include high-quality resources for training on laparoscopic and robotic radical prostatectomy? World J Urol. 2020;38:1195-9.
8. Celentano V, Smart N, McGrath J, Cahill RA, Spinelli A, Obermair A, et al. LAP-VEGaS practice guidelines for reporting of educational videos in laparoscopic surgery: a joint trainers and trainees consensus statement. Ann Surg. 2018;268:920-6.
9. Ferhatoğlu MF, Kartal A, Filiz Aİ, Kılıç A, Aydın MT. Evaluation of the reliability, utility, and quality of the information in sleeve gastrectomy videos shared on open access video sharing platform YouTube. Obes Surg. 2019;29:1477-84.
10. Erdem MN, Karaca S. Evaluating the accuracy and quality of the ınformation in kyphosis videos shared on YouTube. Spine (Phila Pa 1976). 2018;43:E1334-E1339.
11. Kerrey BT, Rinderknecht AS, Geis GL, Nigrovic LE, Mittiga MR, Fine AM, et al. Rapid sequence intubation for pediatric emergency patients: higher frequency of failed attempts and adverse effects found by video review. Ann Emerg Med. 2012;60:251-9.
12. Bochenska K, Milad MP, Santoro N, Twidale E, McGuire J, Ferris D. Instructional video and medical student surgical knot-tying proficiency: randomized controlled trial. JMIR Med Educ. 2018;4:e9.
13. Pratsinis M, Abt D, Müllhaupt G, Langenauer J, Knoll T, Schmid HP, et al. Systematic assessment of information about surgical urinary stone treatment on YouTube. World J Urol. 2021;39:935-42.
14. Kaya HD, Dinç H, Yıldız H, Kızılkaya Beji N. Content analysis and quality evaluation of YouTube videos on Leopold's Maneuvers: a mixed-method study. Education and Information Technologies. 2025:1-20.
15. Pavuloori M, Lin A, Mi M. Tools/instruments for assessing YouTube videos on surgical procedures for patient/consumer health education: a systematic review. Front Public Health. 2025;13:1575801.
16. Manchikanti L, Benyamin RM, Falco FJE, Hirsch JA. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: part 6. Systematic reviews and meta-analyses of observational studies. Pain Physician. 2009:819.
17. Akdemir E, Ateş M, Özcan C, Koç G, Gökçe Mİ, Hamidi N. Educational quality of YouTubeTM videos on laparoscopic radical prostatectomy. J Laparoendosc Adv Surg Tech A. 2025;35:373-8. Epub 2025 Mar 21.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Review
DOI: 10.4274/eajem.galenos.2025.28412
Eurasian J Emerg Med. 2026;25: 117-24

Large Language Models in Emergency Medicine: A Critical Appraisal of Validity, Reproducibility, and Clinical Utility (2020-2025)

Ahmet Aykut, Cem Yıldırım, Ertuğ Günsoy

University of Health Sciences Türkiye, Van Education and Research Hospital, Clinic of Emergency Medicine, Van, Türkiye

Abstract

Recent studies on large language models (LLMs) in emergency medicine (EM)-have expanded rapidly, yet core threats to validity and reproducibility remain under-addressed. We critically synthesized the methods, reporting quality, and clinical relevance of LLM-focused work in emergency care published between January 2020 and April 2025. We conducted a PubMed search and verified journal indexing in the Web of Science (WoS) to restrict screening to EM-relevant studies published in journals indexed under the WoS 'EM' category, excluding editorials that lacked primary or secondary analysis. Two reviewers independently coded protocol availability; prompt transparency; data realism; reference standards; calibration and decision-curve reporting; external validation; and expert benchmarking, resolving discrepancies by consensus. Ninety-one studies met the inclusion criteria; sixty were original investigations. Prompt disclosure was complete in roughly one-third of studies, and real-world clinical data were used less often than synthetic or examination-style vignettes. Calibration, decision-curve analysis, and demonstrations of incremental value over parsimonious clinical baselines were infrequently reported. Expert benchmarking appeared inconsistently across journal strata, and "near-expert" claims often relied on proxy tasks with limited ecological validity. External validation was uncommon, and model/version identifiers were frequently incomplete, undermining reproducibility. Overall, the current LLM literature within this core EM journal corpus is method-lean and report-light: high-level accuracy claims rarely translate into decision-useful evidence. A minimum reporting set—transparent prompts, code, and versioning; calibration; decision-curve analysis; and expert benchmarking on real data—is needed; absent these elements, deployment in time-critical emergency care remains premature.

Keywords: Large language models, emergency informatics, decision support, methodological evaluation, artificial intelligence

Introduction

Large language models (LLMs), such as OpenAI's GPT and Google's Gemini, have rapidly entered clinical discourse not through careful, protocolized integration but largely propelled by enthusiasm and marketing narratives about their capacity to parse unstructured data, aid diagnostics, and streamline documentation across specialties, including emergency medicine (EM) (1,2). Yet the pace of adoption has outstripped the maturation of evaluation science, creating a widening gap between performance claims and decision-useful, reproducible evidence (3).

Unlike traditional clinical tools, LLMs are non-stationary systems. Their behavior evolves due to silent backend updates, architectural shifts, and access-path changes; the same prompt can yield materially different outputs over time—a form of version drift that undermines reproducibility and weakens causal attribution of observed effects (4). Compounding this instability is prompt sensitivity: minor phrasing changes can produce large, unpredictable swings in outputs, challenging any assumption of determinism or reliability in time-critical EM settings (5).

Despite these well-described failure modes, published LLM evaluations in medicine, including those appearing in core EM journals, frequently omit core methodological details. Model/version identifiers, release dates, full prompt templates, and data provenance are inconsistently reported; synthetic prompts, exam-

Corresponding Author: Ahmet Aykut MD, University of Health Sciences Türkiye, Van Education and Research Hospital, Clinic of Emergency Medicine, Van, Türkiye
E-mail: ahmet.aykut@gmail.com **ORCID ID:** orcid.org/0009-0001-3173-8994

Cite this article as: Aykut A, Yıldırım C, Günsoy E. Large language models in emergency medicine: a critical appraisal of validity, reproducibility, and clinical utility (2020-2025). Eurasian J Emerg Med.2026;25: 117-24.

Received: 23.10.2025
Accepted: 20.12.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

style questions, or proxy tasks often substitute for authentic clinical inputs and adjudicated reference standards (3,6). Consequently, reported accuracies are rarely accompanied by calibration, decision-curve analysis (net benefit), or incremental value over parsimonious clinical baselines—elements required to judge whether a tool improves triage, resource allocation, or patient-relevant outcomes under real EM constraints (7).

This review systematically evaluates peer-reviewed LLM studies pertinent to EM, as represented by studies published in journals indexed under the Web of Science (WoS) "emergency medicine" category (SCI-E/ESCI), focusing on methodological transparency, empirical rigor, and clinical grounding. Using a structured framework centered on input fidelity, evaluation strategy, and functional use case, we (1) characterize the evidentiary landscape within this core EM journal corpus, (2) identify recurrent threats to validity and reproducibility, and (3) outline priorities for future research that make EM deployment transparent, calibrated, externally validated, and resilient to technological volatility (3,6,7).

Review Scope and Approach

Scope and Conceptual Approach

We conducted a retrospective, descriptive review of the peer-reviewed literature to evaluate how LLMs [including ChatGPT/Generative Pre-trained Transformer (GPT)-4 and Gemini] have been applied, evaluated, and reported methodologically in EM. The review emphasized three core dimensions aligned with EM workflows: (1) transparency in model use (versioning, access modality, prompt disclosure), (2) empirical rigor (real-world data, expert benchmarking, calibration, and decision-curve analysis), and (3) clinical applicability (task-workflow fit). A structured, three-axis framework—input fidelity, evaluation strategy, and functional use case—guided classification and subsequent comparisons.

Literature Identification and Study Selection

Information Sources and Search Strategy

A structured PubMed search was performed in May 2025 directly via the PubMed web search interface using the Boolean string below to capture EM-relevant LLM studies between 01/01/2020 and 04/30/2025: "ChatGPT" or "(GPT)-3" or "GPT-3.5" or "GPT-4" or "GPT-4.5" or "large language model" or "LLM" or "generative AI" or "generative artificial intelligence" or "transformer-based model" or "foundation model" or "instruction-tuned model" or "GPT" and [2020.01.01 (date-publication): "2025.04.30" (date-publication)]. The search returned 12.125 records prior to de-duplication and journal-scope filtering.

Journal Corpus Definition (WoS Emergency Medicine Category)

To define a core corpus of EM journals and to ensure comparability of scope across included venues, we retained only articles published in journals indexed in WoS under SCI-E or ESCI. WoS was used solely to verify the SCI-E/ESCI indexing status of journals; no literature search was conducted within WoS. A curated list of 57 EM-category journal titles was used for this index-verification filter.

Screening and Eligibility Assessment

Two independent reviewers screened titles and abstracts, and subsequently full texts, against pre-specified criteria. Reasons for exclusion included: (1) passing mention of LLMs without evaluative content, (2) general AI discussion lacking LLM-specific evaluation, and (3) abstract-only or unavailable full text. Disagreements were resolved by consensus after discussion. Following screening, 91 studies were retained for full-text analysis. Operational definitions, decision rules, and study-level coding outputs are provided in Supplementary Table S1.

Data Abstraction and Methodological Appraisal

All 91 studies were assessed using a structured codebook covering five methodological indicators: model identification (model name, release/version, and access pathway), prompt disclosure (full, partial, absent, or ambiguous), use of real clinical data (authentic EM inputs vs. synthetic/proxy tasks), human expert benchmarking (comparison with domain experts where clinical judgment is implicated), and methodological consistency (alignment between stated aims and data/evaluation choices). Articles were independently coded by two reviewers; disagreements were adjudicated by consensus. Because our primary aim was appraisal of transparency and reporting rather than effect estimation, we did not perform a formal risk-of-bias assessment. The full codebook (variable definitions and decision rules) and the per-study coding matrix are provided in Supplementary Table S1.

High-Rigor Subset (Predefined Methodological Benchmark)

Among the 60 studies classified as original research, we applied a secondary filter to identify a high-rigor subset that met all of the following criteria: (1) real-world clinical data, (2) expert-based benchmarking, and (3) no reliance on standardized multiple-choice exams as surrogate evaluations (to mitigate data leakage and poor ecological validity). Seventeen studies met these criteria.

Three-Axis Classification Framework

To map methodological diversity across original investigations, we applied a three-axis schema: input fidelity (real-world EM

data vs. simulated/synthetic content), evaluation strategy (e.g., expert benchmarking, external validation, calibration/decision-curve reporting vs. internal metrics alone), and functional use case (clinical decision support, documentation/communication, education/training, or other EM-relevant tasks). Operational definitions and coding rules appear in Supplementary Table S1. For visualization, the evaluation strategy and the use case were collapsed into broader themes; these groupings were visualized in the alluvial (Sankey) plot (Figure 1).

Positioning Relative to Reporting Guidance

We compared our classification outputs with healthcare MInimum reporting items for CLear Evaluation of Accuracy Reports (MI-CLEAR)-LLM reporting frameworks MI-CLEAR-LLM and Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD)-LLM. While these guidelines prioritize documentation of model metadata and evaluation design, our schema extends them by explicitly encoding functional role and data realism, enabling granular, metric-specific appraisal. A side-by-side summary is presented in Supplementary Table S2.

Journal-Tier Summaries

Original research articles were grouped according to indexing in SCI-E or ESCI. For each stratum, we calculated the proportions reporting full prompt disclosure, use of real clinical data, and human expert benchmarking. Results are summarized in Table 1 with risk differences (SCI-E-ESCI) and 95% confidence intervals (CI) (Newcombe method). Given the small denominators, particularly in ESCI, findings were interpreted cautiously and treated as descriptive. Because several tier-stratified 2×2 tables involved small expected cell counts (<5), large-sample z- or χ^2-approximations were not used for inference.

Exploratory Cross-tabulations

Exploratory Fisher's exact tests were prespecified for the interaction between journal tier and reporting items (full prompt, real clinical data, expert benchmarking), and for the interaction between prompt disclosure (full vs. not-full) and high-rigor status. Two-sided Fisher's exact tests were used, and the corresponding 2×2 tables and p-values are provided in Table 2. Fisher's exact test was selected because it remains valid when expected cell counts are small, unlike χ^2- and z-based procedures that rely on large-sample assumptions. Other planned tests (e.g., real data × expert benchmarking, use case × prompt) required joint counts not available from the final abstraction set and were not performed.

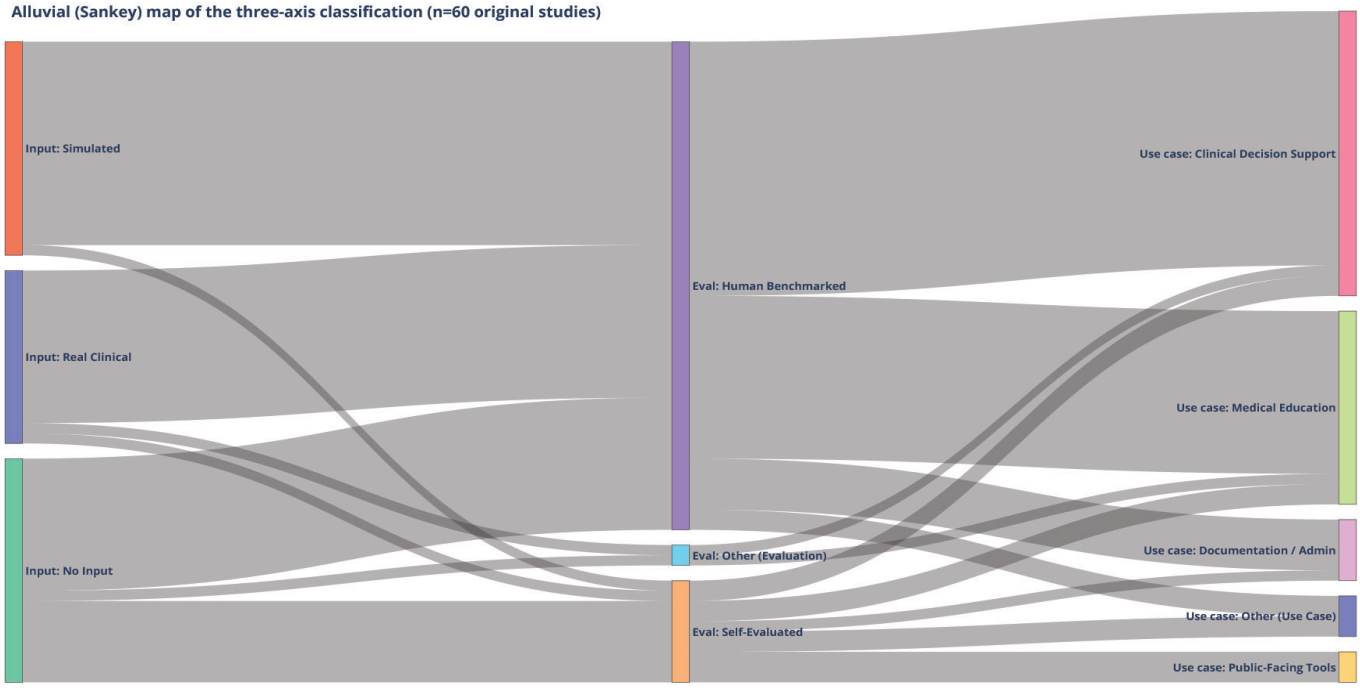


Figure 1. Alluvial (Sankey) map of the three-axis classification of original investigations (n=60)

Nodes represent input fidelity (real clinical / simulated/no input), evaluation strategy (human benchmarked/self-evaluated/other), and functional use cases (clinical decision support/medical education/documentation/admin/public-facing tools/other). Link widths are proportional to the number of studies in each pathway

Evidence Synthesis

Publication Trends and Study Types

Ninety-one peer-reviewed articles met the inclusion criteria, of which sixty were classified as original research. Publications were concentrated between 2023 and 2025, peaking in 2024. Across all included articles, most were published in SCI-E-indexed journals (SCI-E: 80/91; ESCI: 11/91). Among original research articles, 53 appeared in SCI-E-indexed journals and 7 in ESCI-indexed journals. The distribution of publication types by year and journal tier is summarized in Table 3.

Methodological Rigor Subset

Original research articles were evaluated against three a priori criteria: use of real-world clinical data, inclusion of human expert benchmarking, and avoidance of standardized multiple-choice examinations as surrogate tasks. Seventeen studies satisfied all three criteria and were designated as methodologically robust. The remaining forty-three studies did not meet these criteria due to methodological incompleteness or ambiguity (n=32), reliance on simulated data and/or the absence of expert comparison (n=6), or use of examination-style proxies (n=5). Non-robust studies were assigned a primary reason using a mutually exclusive hierarchy, as reflected in the study-level coding (Supplementary Table S1). Counts and definitions are presented in Table 3, with study-level coding in Supplementary Table S1.

Prompt Transparency and Implications for Reproducibility

Prompt reporting was inconsistent across the original research corpus. Only 20 studies (33.3%) disclosed complete prompt text, 36 (60.0%) provided partial or paraphrased examples, 3 (5.0%) did not report prompts, and 1 (1.7%) used ambiguous formatting. Even among the methodologically robust subset, full prompt disclosure was observed in six of the seventeen studies. This limited transparency constrains reproducibility for tasks in which model behavior is highly sensitive to input design, including diagnostic reasoning, triage, and clinical documentation. Tier-stratified prompt disclosure is summarized in Table 1, and the exploratory 2×2 comparison of full prompt disclosure versus high-rigor status is reported in Table 2; study-level prompt coding is provided in Supplementary Table S1.

Journal-tier Summary

Descriptive comparisons between SCI-E and ESCI strata are shown in Table 1. Full prompt disclosure was reported in 18 of 53 SCI-E studies (34.0%) and in 2 of 7 ESCI studies (28.6%), yielding a risk difference of +5.4 percentage points (95% CI: -41.4 to +39.2; Newcombe method). The use of real clinical data was nearly identical across tiers—15/53 (28.3%) in SCI-E and 2/7 (28.6%) in ESCI—corresponding to a risk difference of −0.3 percentage points (95% CI: -46.1 to +33.3 percentage points). Human-expert benchmarking was reported in 17 of 53 SCI-E studies (32.1%) and 3 of 7 ESCI studies (42.9%), yielding a risk difference of -10.8

Table 1. Reporting metrics by journal tier (SCIE vs. ESCI): proportions and risk differences (95% CI)

Metric	SCIE (x/n)	SCIE (%)	ESCI (x/n)	ESCI (%)	Risk difference (SCIE-ESCI), %	95% CI (%, Newcombe)
Full prompt disclosure	18/53	34	2/7	28.6	5.4	(-41.4, 39.2)
Use of real clinical data	15/53	28.3	2/7	28.6	-0.3	(-46.1, 33.3)
Human expert benchmarking	17/53	32.1	3/7	42.9	-10.8	(-53.9, 29.7)

Values are presented as x/n (%) for SCIE and ESCI original research; the risk difference (SCIE-ESCI) and 95% CI (Newcombe method) are also reported. Positive RD favors SCIE; negative RD favors ESCI. Interpret with caution due to the small ESCI denominator. SCIE: Science Citation Index-Expanded, ESCI: Emerging Sources Citation index, CI: Confidence interval

Table 2. Exploratory Fisher's exact tests

Comparison	Group 1 (row 1)	a	b	Group 2 (row 2)	c	d	Odds ratio (Fisher's exact tests)	p (two-sided)
Journal tier × Full prompt disclosure	SCIE (full, not full)	18	35	ESCI (full, not full)	2	5	1.286	1.000
Journal tier × Real clinical data	SCIE (real, not real)	15	38	ESCI (real, not real)	2	5	0.987	1.000
Journal tier × Human expert benchmarking	SCIE (Yes, No)	17	36	ESCI (Yes, No)	3	4	0.630	0.676
Prompt (full vs not-full) × High-rigor (Yes, No)	High-rigor = Yes (full, not full)	6	14	High-rigor = No (full, not full)	11	29	1.130	1.000

Cells follow the 2×2 convention: [(a, b), (c, d)]. For journal-tier comparisons, row 1 =SCIE and row 2 =ESCI; columns indicate feature presence and absence. For the prompt × high-rigor comparison, row 1= High-rigor= Yes, row 2= High-rigor= No; columns are (full, not full). Two-sided Fisher's exact tests p-values are reported, SCIE: Science Citation Index-Expanded, ESCI: Emerging Sources Citation index

percentage points (95% CI: -53.9 to +29.7 percentage points). Overall, CIs were wide and included the null, limiting inference about tier-level differences and highlighting the imprecision resulting from the small ESCI denominator. Accordingly, tier comparisons are interpreted descriptively, and inferential testing relies on Fisher's exact test given the small expected cell counts.

Exploratory Cross-tabulations

Pre-specified exploratory 2×2 tests are reported in Table 2. Two-sided Fisher's exact tests did not identify statistically significant associations between journal tier and full prompt disclosure (odds ratio: 1.286, p=1.000), real clinical data (odds ratio: 0.987, p=1.000), or human expert benchmarking (odds ratio: 0.630, p=0.676). Likewise, the association between full prompt disclosure and inclusion in the high-rigor subset was not significant (odds ratio: 1.130, p=1.000). Planned analyses for real data × expert benchmarking, and use case × prompt transparency could not be conducted because joint counts were not available in the final abstraction; these omissions are noted in Methods 2.8. All contingency tables and p-values are presented in Table 2.

Methodological Structure Across Studies

Using the three-axis framework (input fidelity, evaluation strategy, functional use case), we characterized methodological patterns across the sixty original research articles. Real clinical data were used in 28.3% of studies (17/60), and human expert benchmarking was reported in 33.3% (20/60). Clinical decision support was the most frequent use case, followed by educational and administrative applications (Supplementary Table S1). Overall, the corpus exhibits rapid growth alongside substantial heterogeneity in reporting and design, with a small but identifiable subset meeting more stringent criteria for empirical and clinical rigor. The distribution of studies across the three axes and their interrelationships are visualized in the alluvial (Sankey) plot (Figure 1).

Discussion

Fragile Foundations Amid Rapid Growth

The present synthesis reveals a corpus expanding faster than its evidentiary substrate. Following the release of contemporary LLMs (e.g., GPT-4), publication volume rose sharply; however, reporting quality and clinical grounding have not kept pace within the core EM journal set examined. Much of the literature remains anchored in capability demonstrations rather than decision-useful evidence, a gap reflected in incomplete description of model provenance, insufficient transparency around inputs, and limited assessment of clinical impact. These findings align with prior critiques of methodological shortfalls and checklist non-adherence in medical LLM research. (1-3,6,7).

Version Instability and Prompt Sensitivity as Threats to Reproducibility

LLMs are non-stationary: silent model updates, undocumented architectural changes, and evolving access pathways introduce version drift, whereby identical prompts can yield divergent outputs across time (4,5). Together with prompt sensitivity, defined as large output variability induced by minor input changes, this dynamic undermines reproducibility and complicates

Table 3. Tier totals

Journal tier	Commentary (n)	Original research (n)	Review (n)	Total (n)
ESCI	1	7	3	11
SCIE	27	53	0	80
All	28	60	3	91

Counts are numbers of articles by publication year and Web of Science Emergency Medicine journal tier. "Original research" denotes empiric investigations; "commentary" includes editorials, viewpoints, and letters; "review" comprises narrative or systematic reviews. SCIE: Science Citation-Index Expanded, ESCI: Emerging Sources Citation index. Row totals equal the sum of commentary, original research, and review for each year-tier combination; tier totals and the grand total (SCIE: 80, ESCI: 11, overall n=91) are reported beneath the main table. Figures reflect the curated Emergency Medicine journal set used for study selection and may not generalize beyond this corpus

Table 3. Distribution of studies by year, journal tier, and publication type (2023-2025)

Year	Journal tier	Commentary (n)	Original research (n)	Review (n)	Total (n)
2023	ESCI	1	2	1	4
2023	SCIE	13	10	0	23
2024	ESCI	0	4	2	6
2024	SCIE	10	27	0	37
2025	ESCI	0	1	0	1
2025	SCIE	4	16	0	20

SCIE: Science Citation Index-Expanded, ESCI: Emerging Sources Citation index

longitudinal interpretation. Reproducibility challenges are not unique to LLMs; variability across runs and pipelines has also been documented in other medical AI domains such as deep learning–based image segmentation (8). Studies that omit model/version identifiers, release timing, or access modality produce findings that are difficult to replicate or reconcile with subsequent evaluations in this corpus (3-5).

Transparency Deficits and Their Implications

Prompt disclosure was inconsistent in the included EM-journal corpus: complete prompts were reported in only one-third of original studies and even less frequently in the methodologically robust subset. Absence or partial reporting of prompts constrains replication, especially for tasks in which outputs are highly sensitive to input design (diagnostic reasoning, triage, documentation). These observations are consistent with emerging guidance that positions transparent prompts, versioning, and code availability as prerequisites for credible evaluation (3,6,7).

Methodological Weaknesses in Study Design

A recurrent limitation was the reliance on standardized multiple-choice or examination-style instruments as proxy measures. Such tasks are vulnerable to training-data contamination and fail to capture EMs complexity, thereby risking optimistic but clinically uninformative estimates (6). By contrast, the high-rigor subset combined real clinical inputs with expert benchmarking and avoided using exam proxies, illustrating that empirically grounded LLM evaluation in EM is feasible when the study design is aligned with clinical workflow, as demonstrated in studies published in core EM journals.

Journal-tier Comparisons Interpreted with Caution

Tier-stratified summaries suggested a slightly higher prompt disclosure in SCI-E, no meaningful difference in the use of real clinical data, and no consistent advantage in human expert benchmarking (the ESCI proportion was numerically higher); CIs were wide across all metrics. However, these tier-stratified comparisons are constrained by the small ESCI subgroup (n=7) relative to SCI-E (n=53), which limits precision and precludes robust inference. These patterns argue against tier-level generalizations and should be interpreted as descriptive patterns within the WoS "EM" journal category rather than as field-wide differences. Accordingly, tier patterns are not used to support the study's primary conclusions.

Beyond Checklists: A Structural Lens on Study Quality

Reporting frameworks such as MI-CLEAR-LLM and TRIPOD-LLM have improved transparency, but cannot, by themselves, remedy deeper design deficits (3,7). Our three-axis schema—input fidelity, evaluation strategy, and functional use case—provides a complementary structural framework, highlighting the persistence of synthetic tasks, the infrequency of calibration and decision-curve analyses, and the variability of expert benchmarking, even among studies presented as rigorous within the examined EM-journal corpus. The framework's relationship to existing guidance is detailed in Supplementary Table S2.

Structural Barriers to External Validation

Opaque model internals, undisclosed training data, and proprietary constraints impede error analysis and durable external validation (9-11). Absent auditable versioning and accessible artifacts (prompts/code), even well-designed evaluations risk becoming brittle as platforms evolve. A version-aware, documentation-first approach is therefore integral to any clinically credible assessment strategy for EM-facing use cases, including those evaluated in core EM journals.

Toward a More Reliable Evaluation Paradigm

The field should pivot from leaderboard-style claims of accuracy to process-oriented evaluation that prioritizes version tracking, prompt transparency, calibration, decision-curve analysis, external validation, and expert benchmarking using authentic EM data (1,3,6,7,9). Studies exemplifying these elements already exist, demonstrating the feasibility and value of methodologically disciplined LLM research in EM, as reflected by a subset of rigorously designed studies identified in this journal corpus (12,13). Consolidating such practices will be essential to move from exploratory promise to defensible deployment.

Methodological Considerations

This review was restricted to journals indexed in the WoS "EM" category (SCI-E and ESCI); therefore, our findings should be interpreted as describing the LLM evidence base within this core EM journal corpus, rather than the entire EM literature. Relevant EM-facing LLM studies published in general medicine, critical care, resuscitation, trauma, prehospital/EMS, radiology, and medical informatics venues may have been overlooked. This venue restriction may introduce publication-location (venue) bias, potentially affecting the observed distribution of use cases, evaluation designs, and reporting practices. For example, clinically grounded or implementation-focused studies may be more likely to appear outside EM-category journals, whereas capability demonstrations may cluster differently across venues.

Study-level coding was performed systematically, yet some classifications, particularly along the evaluation-strategy and use-case axes, required interpretive judgment due to inconsistent reporting in source articles. To enable visual synthesis, we collapsed several categories in the alluvial mapping, a simplification that may obscure finer methodological distinctions. We did not

perform a sensitivity analysis using an expanded journal set beyond the WoS EM category; future work should assess the robustness of these patterns by broadening sampling frames (e.g., predefined acute-care-relevant journal lists and/or topic-based retrieval with EM-relevance adjudication). Tier-stratified summaries should also be interpreted cautiously because the ESCI subgroup was small (n=7) relative to SCI-E (n=53), limiting precision and precluding robust inference about tier differences. In addition, some tier-stratified 2×2 tables had expected cell counts below conventional thresholds (e.g., <5), which preclude reliable χ^2- or z-based inference and z-score calculations. We therefore relied on Fisher's exact tests; this reliance contributes to wide CIs and low power to detect differences between tiers. Consequently, these tier-stratified estimates have limited clinical usability and generalizability, and should not be used to draw tier-level inferences. Finally, we did not undertake a longitudinal analysis of temporal drift in model behavior or prompt practices; given ongoing platform evolution, future meta-research should explicitly incorporate time-stamped versioning and repeated-measures designs.

Recommendations and Conclusion

Report the Model-Precisely and Time-Stamped

Responsible evaluation begins with unambiguous model metadata. Every study should document the model's canonical name, release or build identifier, access pathway (e.g., API versus web interface), and the exact date and time of use. Absent these anchors, results are not reproducible in a non-stationary ecosystem and cannot be meaningfully compared across replications or versions. Contemporary guidance (MI-CLEAR-LLM, TRIPOD-LLM) already positions version transparency as a first-order requirement; clinical LLM research must normalize it as standard practice (3,7).

Treat Prompts as Experimental Conditions

Prompts and associated parameters shape outputs as deterministically as any intervention in a clinical experiment. Studies should provide complete prompt text (including system instructions), retry logic, and sampling parameters. Summaries or partial exemplars are insufficient for replication and hinder meta-research on prompt sensitivity. This expectation is both methodologically necessary and aligned with emerging recommendations on reproducible LLM evaluation (3,5,7).

Replace Exam Proxies with Clinical Reality

Standardized multiple-choice items are convenient but of low ecological validity and vulnerable to training-data contamination. Evaluations should be grounded in authentic clinical artifacts — electronic health records, clinician notes, and imaging reports — and should mirror real emergency workflows and constraints. Synthetic benchmarks frequently overstate performance and obscure operational failure modes; they should be used only for preliminary exploration and not presented as evidence of clinical readiness (3,6,12).

Benchmark Against Human Experts

For high-stakes tasks (diagnosis, triage, disposition), human expert comparators and adjudicated reference standards are indispensable. Model-to-model comparisons are insufficient for clinical inference and pose a risk of circular validation. The clinical literature and reporting guidance converge on this point: expert benchmarking is a prerequisite for decision-useful claims (3,7,12,13).

Shift the Emphasis from Point Accuracy to Process Understanding

Leaderboard accuracy is an inadequate proxy for safety or usefulness. Future studies should interrogate reasoning pathways and failure modes, incorporating calibration, decision-curve analysis, error taxonomies, and qualitative audits. Robustness, not isolated point estimates, must become the currency of clinical evaluation (9,10).

Anticipate—and Measure—Drift

Because LLMs evolve over time, validation should include repeated assessments across versions and time points, with explicit cross-version comparisons and time-stamped artifacts. Version drift is no longer hypothetical; it is a documented source of variance that threatens longitudinal interpretability if not prospectively measured (4,7,9,10).

Enforce Minimum Reporting at Publication

Editorial policies should require, as a condition of acceptance, complete model metadata, full prompt disclosure, and clinically grounded evaluation design. Manuscripts lacking these elements are methodologically incomplete regardless of novelty or apparent performance. Reproducibility and transparency —not speed to publication— should define publication worthiness (1,3,7,9).

Conclusion

LLMs hold clear promise for EM, particularly in decision support, documentation, and clinical education. Realizing that promise, however, demands a decisive shift in scientific practice: version-aware reporting, full prompt transparency, expert-anchored evaluation on real clinical data, and routine use of calibration and decision-curve analysis. Within the body of LLM studies published in core EM journals (WoS "EM" category), our synthesis

shows that the current evidence base remains fragmented, with critical gaps in documentation and external validity that limit clinical translation. The path forward is not to celebrate higher scores on synthetic tasks but to establish rigorous, reproducible, and clinically coherent methods. If EM embraces this standard—prioritizing process over performance and accountability over expediency—LLM integration can proceed with the trust and durability required for high-stakes care.

Ethics

Footnotes

Authorship Contributions

Concept: A.A., Design: A.A., Data Collection or Processing: A.A., C.Y., E.G., Analysis or Interpretation: A.A., C.Y., E.G., Literature Search: A.A., Writing: A.A., C.Y., E.G.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Haghani M. Riding the wave of ChatGPT research: an analysis of early-stage scholarly output and the associated authorship anomalies. SSRN. Available from: https://www.ssrn.com/abstract=4553479
2. Karabacak M, Margetis K. Embracing large language models for medical applications: opportunities and challenges. Cureus. 2023;15:e393305.
3. Ko JS, Heo H, Suh CH, Yi J, Shim WH. Adherence of studies on large language models for medical applications published in leading medical journals according to the MI-CLEAR-LLM Checklist. Korean J Radiol. 2025;26:304-12.
4. Chen L, Zaharia M, Zou J. How is ChatGPT's behavior changing over time? arXiv [Preprint]. Available from: http://arxiv.org/abs/2307.09009
5. Liu P, Yuan W, Fu J, Jiang Z, Hayashi H, Neubig G. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. ACM Comput Surv. 2023;55:35.
6. Shool S, Adimi S, Saboori Amleshi R, Bitaraf E, Golpira R, Tara M. A systematic review of large language model (LLM) evaluations in clinical medicine. BMC Med Inform Decis Mak. 2025;25:117.
7. Gallifant J, Afshar M, Ameen S, Aphinyanaphongs Y, Chen S, Cacciamani G, et al. The TRIPOD-LLM statement: a targeted guideline for reporting large language models use. medRxiv. 2024;25:2024.
8. Renard F, Guedria S, Palma N, Vuillerme N. Variability and reproducibility in deep learning for medical image segmentation. Sci Rep. 2020;10:13724.
9. Semmelrock H, Kopeinik S, Theiler D, Ross-Hellauer T, Kowald D. Reproducibility in machine learning-driven research. arXiv [Preprint]. Available from: http://arxiv.org/abs/2307.10320
10. Barberis A, Aerts HJWL, Buffa FM. Robustness and reproducibility for AI learning in biomedical sciences: RENOIR. Sci Rep. 2024;14:1933.
11. Roustan D, Bastardot F. The clinicians' guide to large language models: a general perspective with a focus on hallucinations. Interact J Med Res. 2025;14:e59823.
12. Lee KL, Kessler DA, Caglic I, Kuo YH, Shaida N, Barrett T. Assessing the performance of ChatGPT and Bard/Gemini against radiologists for Prostate Imaging-Reporting and Data System classification based on prostate multiparametric MRI text reports. Br J Radiol. 2025;98:368-74.
13. Sonoda Y, Kurokawa R, Nakamura Y, Kanzawa J, Kurokawa M, Ohizumi Y, et al. Diagnostic performances of GPT-4o, Claude 3 Opus, and Gemini 1.5 Pro in "diagnosis please" cases. Jpn J Radiol. 2024;42:1231-5.

Supplementary Table S1 - Table S2: https://d2v96fxpocvxx.cloudfront.net/580eb5e7-1480-44a6-9404-b8b7446acbcb/content-images/bde585b6-1102-437f-bfa1-2f70f12a1d2b.pdf

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.00878
Eurasian J Emerg Med. 2026;25: 125-34

Quality of Three Chest Compression Techniques During Two-Rescuer Infant CPR: A Randomised Crossover Manikin Study

Mehmet Koşargelir[1], Marek Solecki[2], Halla Kaminska[3], Lukasz Szarpak[4], Başar Cander[5], Bartosz Maj[6], Aneta Czupryniak[6], Wojciech Wieczorek[7]

[1]University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye
[2]LUXMED Group, Clinic of Clinical Research and Development, Warsaw, Poland
[3]LUXMED Group, Clinic of of Children, Warsaw, Poland
[4]The John Paul II Catholic University of Lublin, Technology Transfer Center, Lublin, Poland
[5]Bezmialem Vakıf Universiy Faculty of Medicine, Department of Emergency Medicine, İstanbul, Türkiye
[6]Health Center of Tomaszow Mazowiecki, Clinic of Pediatrics, Health, Tomaszow Mazowiecki, Poland
[7]Medical University of Warsaw Faculty of Medicine, Department of Emergency Medicine, Warsaw, Poland

Abstract

Aim: To evaluate the quality and ergonomic impact of three infant chest compression techniques the two-thumb encircling technique (TTHT), the cross-thumb technique (CTT), and the one-hand open-palm technique (OHT) during two-rescuer infant cardiopulmonary resuscitation (CPR) utilizing a 15:2 compression-to-ventilation ratio.

Materials and Methods: This prospective, randomized, crossover simulation study included 50 registered nurses who performed three 2-minute CPR sequences on an infant manikin, each using one of the three techniques. The primary outcomes were the depth of the compression, the percentage of target-range compressions, and the chest compression fraction (CCF). Secondary outcomes encompassed compression rate, recoil, hand position accuracy, excessive compressions, fatigue, pain, perceived difficulty, and hand slippage. We used repeated-measures statistical models to look at the data.

Results: Both thumb-based methods (TTHT and CTT) yielded significantly deeper compressions, elevated proportions of target-range compressions, increased CCF, and enhanced accuracy in hand positioning in comparison to OHT (all $p<0.01$). TTHT and CTT exhibited similar mechanical performance in all primary outcomes. However, CTT had much less fatigue, hand pain, and perceived difficulty than TTHT (all $p<0.01$). OHT caused shallower compressions, the lowest CCF, more over-depth compressions, and the highest rate of hand slippage.

Conclusion: During two-rescuer infant resuscitation, TTHT and CTT are better than OHT at biomechanical CPR quality. CTT has the same compression quality as TTHT but is more comfortable to use, making it a good choice when thumb-based techniques are feasible. OHT should only be used when it is not possible to wrap the chest. More clinical studies are needed to support these simulation results.

Keywords: Infant cardiopulmonary resuscitation, chest compression techniques, cardiopulmonary resuscitation, simulation study, compression quality, pediatric basic life support

Introduction

Infant cardiopulmonary resuscitation (CPR) is vital in emergency pediatric care and closely linked to survival and neurological outcomes. In infants, cardiac arrest usually results from respiratory failure or asphyxia, not heart disease. Thus, both proper chest compressions and effective ventilation are necessary (1-3). The American Heart Association (AHA) and American Academy of Pediatrics (AAP) recommend a 15:2 compression-to-ventilation ratio when two healthcare providers perform CPR on infants or children. This delivers more rescue breaths and aligns with the asphyxial nature of most pediatric arrests (2-4).

Corresponding Author: Lukasz Szarpak Prof, The John Paul II Catholic University of Lublin, Technology Transfer Center, Lublin, Poland
E-mail: lukasz.szarpak@gmail.com **ORCID ID:** orcid.org/0000-0002-0973-5455

Cite this article as: Koşargelir M, Solecki M, Kaminska H, Szarpak L, Cander B, Maj B, et al. Quality of three chest compression techniques during two-rescuer infant CPR: a randomised crossover manikin study. Eurasian J Emerg Med. 2026;25: 125-34.

Received: 23.11.2025
Accepted: 21.12.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Several steps must be taken to ensure high-quality CPR in infants. The correct compression depth is about one-third of the anterior–posterior chest diameter. Hands need to be placed in the correct position, on the lower third of the sternum, and kept away from the xiphoid process (5-7). The chest must fully recoil after each compression. Interruptions in compressions should be kept to a minimum. The compression rate should be 100-120·min^{-1}. Experimental and clinical evidence indicate that incomplete chest recoil, characterized by frequent and prolonged pauses, reduces coronary perfusion pressure. These factors are associated with poorer resuscitation outcomes (8,9). Therefore, any change in technique that affects depth, rate, or recoil could significantly change how well CPR works.

It is crucial to select the right method for chest compressions. The AHA and AAP now recommend the two-thumb encircling technique (TTHT) as the most effective for two-person infant CPR. This method results in deeper compressions, higher coronary perfusion pressures, and less rescuer fatigue compared to the two-finger technique. Despite this, several alternative methods have been proposed and tested in simulations and observational studies over the past decade. However, concerns remain about compression depth, hand stability, and the risk of over-compression.

Although many manikin studies and registry data show infant CPR is effective, few head-to-head comparisons of these three methods exist in realistic two-rescuer, 15:2 basic life support (BLS) scenarios (10-15). Current research suggests the one-hand technique may, under certain conditions, achieve greater depth than thumb-based methods. However, its effects on chest recoil, hand positioning, and rescuer fatigue remain unclear, especially during extended team efforts (4). Studies also show that ergonomic changes such as using a step stool, the “elbow-lock” position, or rotating compressors, can affect compression quality and reduce fatigue. However, these adjustments have not been widely tested with different infant compression techniques (4).

Due to these gaps, we needed a thorough, well-designed comparison of current infant compression techniques under conditions that reflect recommended practice. We compared the two-thumb encircling, cross-thumb, and one-hand open-palm techniques (OHT) during two-rescuer infant CPR using a 15:2 ratio.

Materials and Methods

Study Design and Oversight

We conducted a prospective, randomized, three-arm crossover study. It compared the quality of infant chest compressions given with three techniques: (1) two-thumb encircling, (2) cross-thumb, and (3) one-hand (open-palm). An infant manikin was used in a high-fidelity pediatric simulation setting for all scenarios. Following modern pediatric resuscitation guidelines, each sequence involved two rescuers giving BLS for two minutes with a 15:2 compression-to-ventilation ratio. Participants were randomly assigned to the order in which they performed the three techniques, using a computer-generated allocation sequence.

The Institutional Review Board of the Polish Society of Disaster Medicine reviewed and approved the protocol (approval no: 14/02/2024, date: 14.02.2024). All study procedures adhered to the Declaration of Helsinki and relevant national regulations governing research with human subjects. Volunteers received written and verbal information and provided written consent before participating. The study followed EQUATOR-aligned standards for simulation research. This included specifying prospective endpoints, detailing the simulation model and parameters, and documenting the data structure and analysis workflow.

Participants and Sample Size

The study population included 50 registered nurses. All participants were enrolled in scheduled BLS training courses, taught by AHA–certified instructors. Each participant was actively working in clinical practice and regularly cared for pediatric or neonatal patients. Eligibility required current employment as a nurse, prior completion of at least one BLS course, and willingness to participate in simulation research. Nurses who reported musculoskeletal conditions that made chest compressions unsafe were excluded. Participation was voluntary with no financial incentives. Every participant was assigned a unique ID number. All nurses performed each of the three compression techniques, using a within-subject, randomized three-period crossover design. This resulted in 150 resuscitation attempts (50 participants × 3 techniques).

Simulation Protocol

For each study condition, participants performed a standard infant BLS scenario using a dummy. Each 120-second sequence required nurses to administer chest compressions and ventilations continuously. According to the latest pediatric BLS guidelines, the compression–ventilation ratio was 15:2. Each participant delivered 15 compressions, followed by two ventilations, and repeated this cycle for the full 2 minutes.

The order of the three compression techniques was randomized in a crossover manner. ResearchRandomizer (Randomizer.org; Figure 1) was used to create a unique sequence of techniques for each participant. Allocations were made before enrollment.

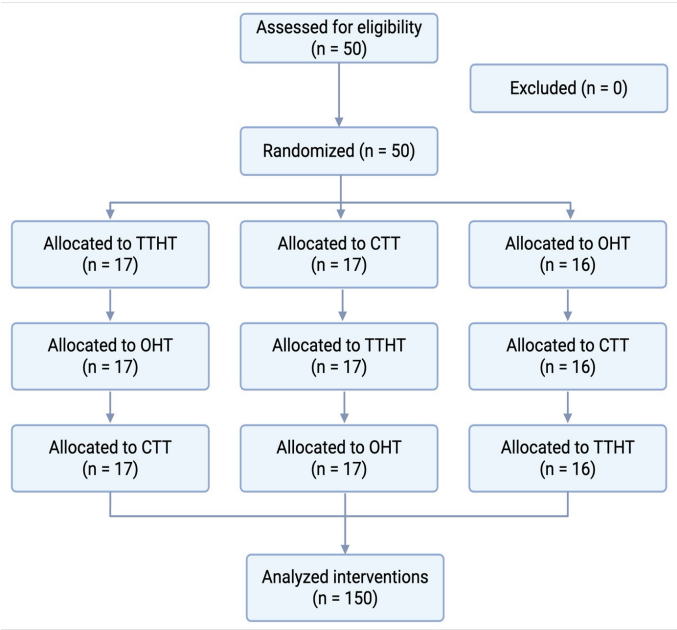


Figure 1. Randomization flow chart
TTHT: Two-thumb encircling technique, CTT: Cross-thumb technique, OHT: Open-palm technique

Participants were assigned to the following available sequence by numerical order. Each nurse performed all three techniques in a different order. This helped reduce learning fatigue and the effects of menstrual periods.

Before the event began, instructors asked participants to keep a compression rate between 100 and 120 per minute. They did not use a metronome. Instead, participants were told to "compress fast and regularly" to allow for natural variation. A second rescuer, also trained in BLS, performed bag-mask ventilation. This person gave two chest-rise breaths after each set of 15 compressions. This is similar to real-life resuscitations, where short breaks occur between compressions.

To prevent fatigue and minimize the mixing of techniques, participants rested for at least 30 minutes between sequences. They took breaks in the simulation room. However, they did not practice compressions or receive extra coaching. This approach ensured that subsequent attempts were based on random technique allocation, not ongoing instruction or warm-up. Before data collection, an AHA-certified instructor demonstrated each technique to all participants. There was a brief practice session. Instructors did not give corrective feedback during recorded attempts.

Three chest compression techniques were used during the study:

- **TTHT:** The rescuer stood at the baby's feet. Both thumbs pointed up, side by side, on the lower third of the sternum. The fingers wrapped around the thorax to support the back. To compress, both thumbs pressed straight down while the hands encircled the chest. This allowed for full chest recoil after each compression.
- **Cross-thumb technique (CTT):** The rescuer stood beside the manikin. Both thumbs were on the lower half of the sternum, with one crossing over the other at the midline. The fingers of both hands wrapped around the thorax for stability. The overlapping thumbs applied vertical pressure. The focus was on vertical compression, full release between compressions, and no lateral movement.
- **One-hand (open-palm) technique:** The rescuer stood next to the mannequin. The heel of one hand was on the lower sternum. Fingers were held up to avoid the ribs and abdomen. The second hand could rest on the forearm or chest for support, but added no force. Compressions were given using the heel of the hand, pressing straight down to about one-third of the chest depth. The hand fully recoiled after each compression (Figure 2).



Figure 2. Chest compression techniques used during two-rescuer infant cardiopulmonary resuscitation (CPR) with a 15:2 compression-to-ventilation ratio: (A) two-thumb encircling technique (TTHT), (B) cross-thumb technique (CTT), and (C) one-hand open-palm technique (OHT)

For all techniques, participants were instructed to minimize interruptions, avoid leaning between compressions, and maintain their focus on maintaining the correct rate, depth, and 15:2 sequence. During the experimental phase, a different neonatal simulator (Laerdal Medical SimBaby, Laerdal Medical, Stavanger, Norway) was used to ensure standardization of resuscitation.

Outcomes

All outcomes were predetermined and based on objective measurements obtained from the manikin, supplemented by concise post-scenario self-reports from participants. To clarify the outcomes, they were divided into two groups: primary measures of the quality of chest compressions and secondary measures of the mechanics of compressions, hemodynamic surrogates, ergonomics, and technical performance.

Primary Outcomes

The average compression depth, measured in millimeters, was the most critical sign of compression quality. The manikin recorded the depth of each compression for every 2-minute sequence. The average value for that sequence was used in the analysis. Since guidelines recommend that compressions should be at least one-third of the anterior-posterior chest diameter in babies, we also counted how many were within the target range of 40-50 mm. This percentage indicates the frequency with which a rescuer achieved a clinically acceptable depth, not just the average. The chest compression fraction (CCF) was the third primary outcome. It shows how much of the 2 minutes was spent giving compressions. Lower values indicate more time spent in pauses, for example, during ventilations or repositioning.

Secondary Outcomes

We initially looked at the compression rate — the number of compressions per minute averaged over two minutes—to understand how chest compression's function. Additional variables assessed compression technique performance. Full chest recoil was the proportion of compressions where the chest returned to neutral without tilting. According to pediatric BLS standards, the correct hand posture was the proportion of compressions with the thumbs or heel centered on the bottom half of the sternum. We tracked the percentage of "over-depth compressions," or compressions deeper than 50 mm, to assess if the force was excessive. Participants rated their fatigue and hand pain on a 0-10 visual analogue scale immediately after each sequence and the perceived difficulty of the technique on a 5-point Likert scale (1 =very easy to 5 =very difficult) to assess the ergonomic burden of each technique. The nurses' subjective judgments were used to estimate each method's physical and technical demands. Finally, we included measures of safety and technical strength. The hand slippage count revealed how often the compressing hand or thumbs left the sternum during a sequence. This variable had a Poisson-like count distribution.

Sample Size

Based on randomized crossover trials of baby chest compression techniques on manikins (10,16), the sample size was chosen. These investigations showed mean compression depth variations of 2-3 mm and within-subject standard deviations (SDs) of 2.5 mm between approaches, indicating a medium effect magnitude. The primary endpoint for sample size calculation was compression depth (mm).

For a one-way repeated-measures ANOVA with three levels (TTHT, CTT, OHT), $\alpha = 0.05$, 80% power, and $f = 0.25$, G* Power 3.1 suggests 40 participants are sufficient. We planned to enroll 44 individuals to cover dropouts and missing data. Fifty participants completed all three approaches once recruitment exceeded this minimum aim to increase estimate accuracy and secondary outcome power. This final sample size has at least 80% power to detect clinically relevant differences in compression depth and sufficient power to detect secondary outcomes (compression rate and rescuer fatigue).

Statistical Analysis

Statistics were based on each participant's three resuscitation attempts (one for each procedure). Intra-subject variables included technique (TTHT, CTT, OHT). The primary analyses used one-way repeated-measures models with technique as a fixed effect and a random intercept for individuals, similar to repeated-measures ANOVA for this balanced design. The overall impact of the strategy on each outcome was tested using F-tests. Pairwise comparisons between approaches were performed using estimated marginal means, with Holm-Bonferroni adjustment for multiple testing when the omnibus test was significant.

Residual plots, Shapiro-Wilk, and Levene's tests were used to assess normality and homogeneity of variance. Greenhouse-Geisser corrected p-values were reported when Mauchly's test indicated non-sphericity. For non-normal or ordinal outcomes such as fatigue, pain, perceived difficulty, and hand slippage frequency, we used nonparametric Friedman tests and Wilcoxon signed-rank tests with Benjamini–Hochberg correction as sensitivity analyses. These yielded the same statistical significance as the primary models and are not discussed. For symmetric continuous data, the mean ± SD is used, whereas skewed variables are summarized as the median interquartile

range. Effect sizes are reported using partial eta-squared (η^2p) for omnibus testing and Cohen's d for significant pairwise contrasts. All tests were two-sided and had a significance level of $\alpha=0.05$.

All statistical analyses were performed using Python 3.11 (Python Software Foundation, Wilmington, DE, USA) with Pandas, NumPy, and SciPy/statsmodels libraries for data management and modeling, and Matplotlib for figure creation.

Results

Participants and Dataset

Fifty registered nurses working in acute care settings participated in the study, and all of them adhered to the protocol. Each participant executed three resuscitation sequences, one utilizing each chest compression technique, yielding 150 analyzable sequences (50 per technique). There were no missing or excluded observations, and no participant withdrew or deviated from the protocol. This means that all recorded sequences were kept for the final analyses. At the time of the study, all the nurses were undergoing a BLS training session certified by the AHA. They had all already taken at least one accredited BLS course. The research component was optional and separate from the course assessment, and it did not affect completion of the course or certification status.

Chest Compression Quality

Table 1 presents descriptive statistics for all compression variables. There was a big difference in the compression rate between the two methods ($F=3.70$, $p=0.027$). TTHT and CTT had almost the same rates, but OHT had a slightly lower rate. In pairwise tests, TTHT and CTT were different from OHT ($p=0.019$ and $p=0.040$, respectively), but TTHT and CTT were not ($p=0.661$; Figure 3).

The CCF was high in all groups, but it varied depending on the method ($F=9.34$, $p<0.001$). CTT produced the highest CCF, TTHT produced an intermediate value, and OHT produced the lowest. There was no difference in CCF between the two thumb-based techniques ($p=0.085$), but both were better than OHT (TTHT vs OHT $p=0.008$; CTT vs OHT $p<0.001$).

Table 1. Chest compression outcomes

Outcome	TTHT	CTT	OHT	TTHT vs. CTT	TTHT vs. OHT	CTT vs. OHT	ANOVA
Compression rate	112 (110-113)	111.0 (110-113.5)	109 (107-110)	0.661	0.019	0.040	F (2.147)=3.70; p=0.027
	112 (5.0)	111.6 (4.54)	109.4 (6.1)				
Chest compression fraction	0.712 (0.706-0.720)	0.720 (0.715-0.728)	0.702 (0.691-0.708)	0.085	0.008	<0.001	F (2.147)=9.34; p<0.001
	0.712 (0.021)	0.720 (0.025)	0.701 (0.023)				
Compression depth	42.64 (41.86-43.27)	41,18 (40.38-42.22)	39.49 (37.52- 40.68)	0.074	<0.001	0.005	F (2.147)=10.97; p=0,000036
	42.23 (2.42)	31.38 (2.33)	39.42 (4.15)				
Percentage of compression with target depth range 40-50 mm	52.3 (44.-56.6) %	43.1 (36.3-47.4)%	35.0 (35-35) %	0.051	<0.001	<0.001	F (2.147)=21.59; p<0.001
	53.2 (15.2) %	47.4 (14.0)%	37.1 (5.7) %				
Percentage of compressions with full chest recoil	96.6 (96.2-97.2)	97 (96.8-97.2)	95 (94.45-95.30)	0.232	<0.001	<0.001	F (2.147)=40.16; p<0.001
	96.69 (1.29)%	96.98 (1.14)%	94.75 (1.59)%				
Correct hand placement	99.0 (98.9-99.3)	98.9 (98.7-99.1)	97.2 (96.5-97.4)	0.192	<0.001	<0.001	F (2.147)=88.62; p<0.001
	99.1 (0.60)%	98.9 (0.54)%	97.2 (1.10) %				
Percentage to deep over 50mm	2.4 (1.6-2.8)	2.2 (1.5-2.5)	3.35 (3.0-3.9)	0.454	<0.001	<0.001	F (2.147)=13.18; p<0.001
	2.24 (1.39)%	2.05 (1.19)%	3.39 (1.64) %				
Rescuer fatigue	4.3 (4.0-4.7)	3.25 (2.8-3.5)	4.6 (4.2-5.05)	<0.001	0.034	<0.001	F (2.147)=31.86; p<0.001
	4.25 (0.93)	3.19 (0.82)	4.70 (1.13)				
Rescuer hand pain	3.5 (3.3- 3.75)	3.1 (2.8-3.4)	4.3 (3.95-4.6)	0.002	<0.001	<0.001	F (2.147)=33.75; p<0.001
	3.54 (0.67)	3.06 (0.83)	4.31 (0.80)				
Technique difficulty	2.5 (2.25-2.70)	2.2 (2.0-2.35)	2.9 (2.6-3.25)	0.017	<0.001	<0.001	F (2.147)=18.26; p<0.001
	2.48 (0.55)	2.20 (0.61)	2.97 (0.78)				
Hand slippage count	0 (0-1)	0 (0-1)	1 (1-1)	0.863	<0.001	<0.001	F (2.147)=10.30; p<0.001
	0.50 (0.58)	0.52 (0.58)	1.02 (0.77)				

TTHT: Techniques the two-thumb encircling technique, CTT: Cross-thumb technique, OHT: One-hand open-palm technique

Compression depth varied by technique (F =10.97, p<0.001). TTHT and CTT produced deeper compressions than OHT (TTHT vs OHT p<0.001; CTT vs OHT p=0.005); TTHT and CTT did not differ significantly (p=0.074).

There was a significant difference between the techniques in the percentage of compressions within the target depth range (40-50 mm) (F =21.59, p<0.001). Both thumb-based methods achieved a higher percentage of target-depth compressions than OHT (all p<0.001). The only difference between TTHT and CTT was very small (p=0.051). The proportion of excessive compressions (>50 mm) was minimal but markedly greater with OHT compared to TTHT or CTT (F =13.18, p<0.001; both comparisons vs OHT p<0.001); TTHT and CTT exhibited no significant difference (p=0.454).

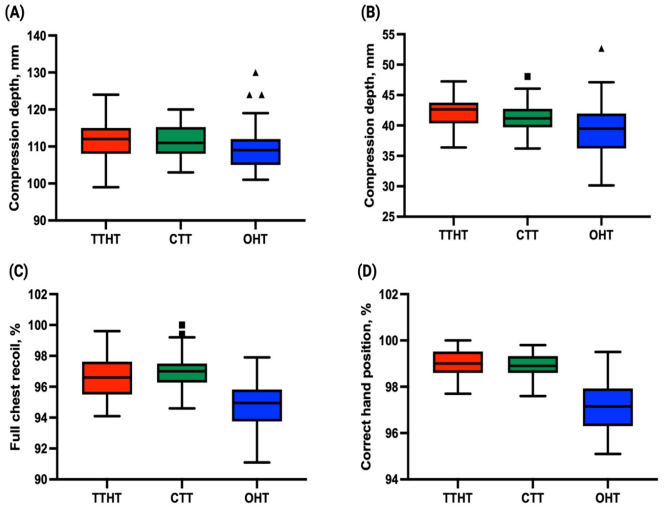


Figure 3. Comparison of chest compression quality parameters between the three infant chest compression techniques: two-thumb encircling technique (TTHT), cross-thumb technique (CTT), and one-hand open-palm technique (OHT). Data are presented as box-and-whisker plots showing median, interquartile range, and minimum/maximum values

Table 1 shows that all techniques achieved nearly full chest recoil, but OHT had slightly less. The technique had a strong effect (F =40.16, p<0.001): TTHT and CTT were similar (p=0.232), both of which were higher than OHT (both p<0.001).

All groups did a great job of placing their hands correctly on the lower half of the sternum; however, the thumb-based methods were more effective (F =88.62, p<0.001). TTHT and CTT were not significantly different from each other (p=0.192), but both were more accurate than OHT (both p<0.001).

Ergonomic Outcomes and Hand Stability

Self-reported fatigue differed by technique (F =31.86, p<0.001; Table 1). CTT had the lowest scores, TTHT was intermediate, and OHT had the highest. CTT caused less fatigue than TTHT and OHT (both p<0.001), and TTHT less than OHT (p=0.034; Figure 4).

Hand pain exhibited a comparable trend (F =33.75, p<0.001): CTT < TTHT < OHT. Pairwise comparisons revealed reduced pain with CTT compared to TTHT (p=0.002) and OHT (p<0.001), as well as diminished pain with TTHT relative to OHT (p<0.001).

The perceived difficulty, rated on a 1-5 scale, also differed by technique (F =18.26, p<0.001). CTT was the easiest, TTHT was a little harder, and OHT was the hardest. The ratings for CTT were lower than those for TTHT and OHT (p=0.017 and p<0.001, respectively), and the ratings for TTHT were lower than those for OHT (p<0.001).

Lastly, the number of times a hand slipped per attempt was low, but technique had a significant effect on it (F =10.30, p<0.001). The number of slippages was similar for TTHT and CTT (p=0.863) but higher for OHT. Both thumb-based techniques had significantly fewer slippages than OHT (both p<0.001).

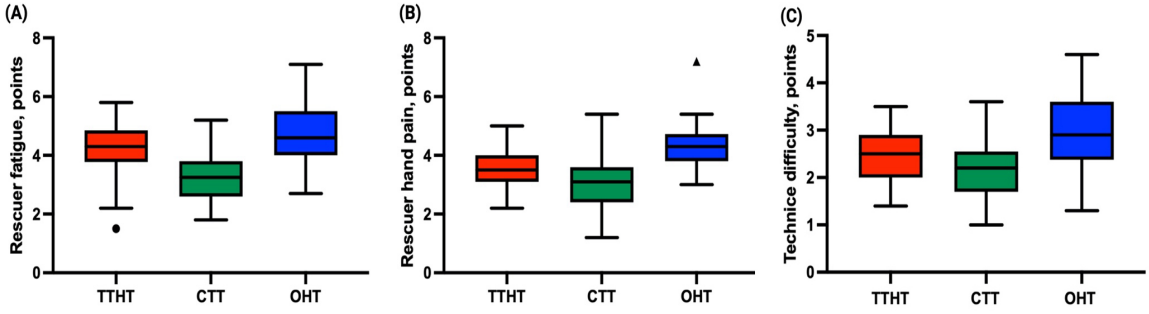


Figure 4. Comparison of ergonomic and subjective outcomes associated with the three infant chest compression techniques: two-thumb encircling technique (TTHT), cross-thumb technique (CTT), and one-hand open-palm technique (OHT). Data are presented as box-and-whisker plots showing median, interquartile range, and minimum/maximum values

Discussion

This randomized crossover simulation study evaluated three infant chest compression techniques –two-thumb encircling hands, cross-thumb, and one-hand open-palm chest– administered by nurses during 2-minute, two-rescuer CPR with a 15:2 ratio. In all core biomechanical outcomes, both thumb-based techniques produced better compressions than OHT. CTT, on the other hand, had similar mechanical performance but much better ergonomic profiles.

Chest Compression Performance and Alignment with Current Guidelines

Our findings support existing resuscitation guidelines, which favor TTHT for baby compression, especially with two rescuers. In our study, TTHT and CTT yielded comparable compression rates and CCF, both considerably higher than OHT. The average TTHT and CTT compression depth was 41-42 mm. OHT made shallower compressions and had the lowest target range compression rate of 40–50 mm (13,14).

These findings confirm earlier manikin and animal studies that TTHT provides greater depth, more consistent compressions, and higher coronary perfusion pressures than the two-finger technique (4,15). Recently, Solecki et al. (16) completed a meta-analysis and narrative review. Due to its better depth accuracy than lateral or two-finger approaches, TTHT is still the best way to perform CPR on babies and neonates. We also found that TTHT beats OHT in a 2-rescuer, 15:2 scenario that follows AHA/AAP baby CPR standards.

The percentage of over-deep compressions (>50 mm) was low across all procedures, except OHT. This shows that providers may overshoot when trying to compensate for one-handed depth. OHT also has the lowest target range compressions, thus it does not offer the best depth or safety. These findings imply that OHT should not be routinely employed when TTHT or CTT are practicable, even though the most recent pediatric BLS recommendations recommend the one-hand technique when the chest cannot be ringed (17).

Cross-thumb Technique Versus Two-Thumb Encircling

A significant contribution of this study is the direct comparison between TTHT and the more recent CTT in a two-rescuer context. We discovered that CTT attained compression rates, CCF, depths, target-depth proportions, and over-depth rates that were statistically indistinguishable from those of TTHT. These results are very similar to those found by Joyner et al. (9) in their randomized manikin study. They found that CTT and TTHT had similar depth and guideline-compliant compression rates, both of which were much better than the two-finger technique (4). Recent comparative data also demonstrate that CTT provides comparable or marginally enhanced mechanical performance while enabling the rescuer to maintain a lateral position, which could be beneficial in confined spaces.

Our findings reveal a significant distinction: despite the mechanical equivalence of TTHT and CTT, CTT was consistently associated with reduced fatigue, diminished hand pain, lower perceived difficulty, and no increase in hand slippage. This ergonomic advantage is clinically significant, as rescuer fatigue is known to degrade CPR quality over time and may be especially detrimental during extended pediatric resuscitation. CTT may be a good alternative to TTHT because it offers both high-quality compressions and reduced physical strain. This is especially true when the compressor must remain next to the patient (for example, in transport incubators, crowded ED bays, or limited workspace around an infant bed).

Nonetheless, it is crucial to underscore that the current evidence for CTT, including our own, is limited to simulation studies. There are no clinical outcome data available to show that CTT is better than or even as satisfactory as TTHT in real cases of infant cardiac arrest. Until such evidence surfaces, TTHT should remain the standard technique endorsed by guidelines, while CTT should be regarded as a promising adjunct or alternative in particular circumstances.

One-hand Technique in the Context of Emerging Evidence

Recent research has rekindled interest in one-handed techniques for infant CPR. Smereka et al.(18) found that the one-hand open-palm technique (OPT) on an infant manikin produced greater depth and was rated as easier than standard two-finger or two-thumb methods. However, the absolute depths still did not consistently meet AHA targets. In a two-rescuer 15:2 scenario, however, our results show that OHT is inferior to TTHT and CTT in several ways: it has shallower compressions, a lower CCF, a significantly lower percentage of target-depth compressions, and a higher percentage of over-deep compressions.

This apparent discrepancy likely stems from differences in how the studies were designed and what they examined. The OPT studies focused on single-rescuer CPR and primarily compared OHT with the two-finger technique, which is commonly acknowledged as inadequate due to its limitations in depth and fatigue. In that context, OHT might be a better option. In our study, OHT was compared to two highly effective thumb-based methods in a two-rescuer scenario, revealing its relative deficiencies. These results support the cautious stance of recent guideline updates, which position the one-hand technique as an option when the provider cannot adequately encircle the chest, rather than as a first-line approach.

Ergonomics and Technical Stability

The ergonomic outcomes in our study provide further insight into the practical trade-offs among techniques. CTT had the lowest scores for fatigue and pain, as well as the lowest perceived difficulty. TTHT was in the middle, and OHT always scored the worst on these measures (13,19).

These results are similar to earlier data, which showed that two-thumb techniques require less rigorous work than two-finger compressions, both in terms of subjective ratings and physiological measures (20). There weren't many instances when hands slipped, but there were significantly more with OHT than with either of the thumb-based methods.

This could be due to the smaller contact area and less stable leverage of the heel of one hand on a small chest. From a patient-safety standpoint, heightened slippage may result in more frequent off-target compressions, especially on compliant infant ribs or in the presence of fluids or gel, although this remains conjectural in the absence of clinical evidence. Our results suggest that OHT may be less forgiving in terms of technical stability, especially for providers with limited pediatric experience, because there are more over-deep compressions.

Clinical and Educational Implications

Our data collectively substantiate multiple pragmatic conclusions for clinical practice and education. First, they support the guidelines that state TTHT should be the standard method for infant CPR when two rescuers are available. Second, they suggest that CTT can be safely regarded as a distinct thumb-based technique that functions mechanically just as well but is more comfortable to use. For teams that frequently revive babies in settings where the compressor must remain near the patient, it may be beneficial to incorporate CTT into local training programs, provided that students also learn TTHT and are aware that it remains the standard (11,21). Third, our results suggest that OHT should not be used when high-quality thumb-based compressions are possible. OHT is useful when the chest can't be encircled, such as with larger babies, unusual body types, or certain situations where the baby can't move. However, our data show that for normal-sized babies and in controlled settings, OHT leads to shallower compressions, less time spent in compression, more technical errors, and increased workload for the rescuer. This nuance may be especially significant as recent consensus documents broaden the application of one-hand techniques in pediatric BLS (22). From an educational standpoint, the substantial influence of technique on both objective quality metrics and subjective workload underscores the imperative of deliberate practice coupled with feedback (1). Training that utilizes simulations to teach students various techniques, focusing on depth, recoil, and minimizing pauses, may help them adjust to different patient sizes, environments, and team compositions without compromising the quality of their care.

Study Limitations

Strengths and Limitations

This study has several strengths. The randomized crossover design, wherein each participant acts as their control, diminishes inter-subject variability and facilitates significant comparison of techniques under uniform conditions. We used a two-rescuer, 15:2 scenario that followed current pediatric guidelines. We also used a group of nurses who had recently completed BLS training and were fairly similar, which makes the results more reliable. There was no missing data in any of the sequences, so the dataset was complete for analysis.

However, significant limitations must be recognized. First, this was a simulation study conducted on an infant manikin; the degree to which observed differences correlate with clinically significant outcomes (return of spontaneous circulation, survival, neurological status) in actual infants remains uncertain. Second, our participants were practicing nurses from a single institution, which may not encompass the full spectrum of pediatric resuscitation experience. Third, we looked at 2-minute bouts of CPR. Rescuer fatigue and technique degradation may differ over longer periods or during more complex resuscitations that involve multiple tasks. Fourth, we examined a single configuration for each technique; alternative configurations, including over-the-head TTHT or adjusted hand positions for CTT, were not investigated. Finally, making multiple comparisons across various endpoints increases the likelihood of committing a Type I error. However, the consistency and magnitude of the effects we observed—especially for OHT versus thumb-based techniques—make it unlikely that our main conclusions are merely random.

Future Directions

Subsequent research must extend beyond the use of manikins to assess these techniques in authentic environments, preferably through multicenter observational registries or pragmatic trials that measure CPR quality metrics during actual infant resuscitations. The incorporation of high-fidelity manikins that can estimate hemodynamic surrogates alongside real-time feedback devices may elucidate whether the ergonomic benefits of CTT result in enduring quality during prolonged resuscitations. Moreover, research contrasting these methodologies in solo-rescuer contexts, during transit, and in restricted settings would enhance the precision of their ideal applications. Finally, cost-effectiveness and learning-curve analyses could help determine

the most effective way to incorporate CTT and OHT into pediatric life support courses without compromising important skills in TTHT.

Conclusion

In this randomized crossover simulation of two-rescuer infant CPR, both TTHT and CTT yielded superior chest compression quality compared to the one-hand OHT, demonstrating increased depth, enhanced target-range compressions, and improved recoil and hand positioning. CTT had similar mechanical performance to TTHT, but it caused less fatigue and pain in the hands and was perceived as easier, suggesting that it could be a practical and more comfortable alternative when thumb-based methods are possible. OHT should still be an option when the chest can't be encircled. Further clinical studies are needed to determine whether the benefits of simulations translate into improved patient outcomes.

Ethics

Ethics Committee Approval: The Institutional Review Board of the Polish Society of Disaster Medicine reviewed and approved the protocol (approval no. 14/02/2024, date: 14.02.2024). All study procedures adhered to the Declaration of Helsinki and relevant national regulations governing research with human subjects. Volunteers

Informed Consent: Volunteers received written and verbal information and provided written consent before participating.

Footnotes

Authorship Contributions

Surgical and Medical Practices: M.S., Concept: M.S., Design: M.S., Data Collection or Processing: M.S., H.K., L.S., B.M., W.W., Analysis or Interpretation: M.K., M.S., L.S., B.M., W.W., Literature Search: M.K., M.S., H.K., L.S., B.C., B.M., A.C., W.W., Writing: M.K., M.S., H.K., L.S., B.C., B.M., A.C., W.W.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Marino BS, Tabbutt S, MacLaren G, Hazinski MF, Adatia I, Atkins DL, et al. Cardiopulmonary resuscitation in infants and children with cardiac disease: a scientific statement from the american Heart Association. Circulation. 2018;137:e691-e82.
2. Hazinski MF, Nadkarni VM, Hickey RW, O'Connor R, Becker LB, Zaritsky A. Major changes in the 2005 AHA guidelines for CPR and ECC: reaching the tipping point for change. Circulation. 2005;112(24 Suppl):IV206-11.
3. Sawyer T, McBride ME, Ades A, Kapadia VS, Leone TA, Lakshminrusimha S, et al. Considerations on the use of neonatal and pediatric resuscitation guidelines for hospitalized neonates and infants: on behalf of the American Heart Association Emergency Cardiovascular Care Committee and the American Academy of Pediatrics. Pediatrics. 2023;153:e2023064681.
4. Joyner BL Jr, Dewan M, Bavare A, de Caen A, DiMaria K, Donofrio-Odmann J, et al. Part 6: Pediatric basic life support: 2025 American Heart Association and American Academy of Pediatrics Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation. 2025;152(16 Suppl 2):S424-S47.
5. Lee HC, Strand ML, Finan E, Illuzzi J, Kamath-Rayne BD, Kapadia V, et al. Part 5: Neonatal resuscitation: 2025 American Heart Association and American Academy of Pediatrics Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation. 2025;152(16 Suppl 2):S385-S423.
6. Lee JE, Lee J, Oh J, Park CH, Kang H, Lim TH, et al. Comparison of two-thumb encircling and two-finger technique during infant cardiopulmonary resuscitation with single rescuer in simulation studies: a systematic review and meta-analysis. Medicine (Baltimore). 2019;98:e17853.
7. Solecki M, Kietlińska M, Pruc M, Nowak-Starz G, Szarpak L, Cander B. Effectiveness and quality of infant and neonatal cardiopulmonary resuscitation: a meta-analysis comparing the over-the-head and lateral positions. Disaster Emerg Med J. 2025;10:195-202.
8. Smereka J, Szarpak L, Rodríguez-Núñez A, Ladny JR, Leung S, Ruetzler K. A randomized comparison of three chest compression techniques and associated hemodynamic effect during infant CPR: A randomized manikin study. Am J Emerg Med. 2017;35:1420-5.
9. Jeon W, Kim J, Ko Y, Lee J. New chest compression method in infant resuscitation: cross thumb technique. PLoS One. 2022;17:e0271636.
10. Smereka J, Szarpak L, Smereka A, Leung S, Ruetzler K. Evaluation of new two-thumb chest compression technique for infant CPR performed by novice physicians. A randomized, crossover, manikin trial. Am J Emerg Med. 2017;35:604-9.
11. Smereka J, Madziala M, Szarpak L. Comparison of two infant chest compression techniques during simulated newborn cardiopulmonary resuscitation performed by a single rescuer: a randomized, crossover multicenter trial. Cardiol J. 2019;26:761-8.
12. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med. 2019;38:2074-102.
13. Cioccari G, Sica da Rocha T, Piva JP. Two-thumb technique is superior to two-finger technique in cardiopulmonary resuscitation of simulated out-of-hospital cardiac arrest in infants. J Am Heart Assoc. 2021;10:e018050.
14. Chang CY, Hou YT, Chien YJ, Chen YL, Lin PC, Chen CS, et al. Two-thumb or two-finger technique in infant cardiopulmonary resuscitation by a single rescuer? A meta-analysis with GOSH analysis. Int J Environ Res Public Health. 2020;17:E5214.
15. Del Rios M, Bartos JA, Panchal AR, Atkins DL, Cabañas JG, Cao D, et al. Part 1: Executive summary: 2025 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation. 2025;152(16 Suppl 2):S284-S312.
16. Solecki M, Tomaszewska M, Pruc M, Myga-Nowak M, Wieczorek W, Katipoglu B, et al. Evaluating novel chest compression technique in infant CPR: enhancing efficacy and reducing rescuer fatigue in single-rescuer scenarios. Children (Basel). 2025;12:346.
17. O'Connell KJ, Sandler A, Dutta A, Ahmed R, Neubrand T, Myers S, et al. The effect of hand position on chest compression quality during CPR in young children: findings from the videography in pediatric resuscitation (VIPER) collaborative. Resuscitation. 2023;185:109741.
18. Smereka J, Kaminska H, Wieczorek W, Dąbrowski M, Ładny JR, Ruetzler K, et al. Which position should we take during newborn resuscitation? A prospective, randomised, multicentre simulation trial. Kardiol Pol. 2018;76:980-6.

19. Smereka J, Iskrzycki Ł, Makomaska-Szaroszyk E, Bielski K, Frass M, Robak O, et al. The effect of chest compression frequency on the quality of resuscitation by lifeguards. A prospective randomized crossover multicenter simulation trial. Cardiol J. 2019;26:769-76.
20. Millin MG, Bogumil D, Fishe JN, Burke RV. Comparing the two-finger versus two-thumb technique for single person infant CPR: a systematic review and meta-analysis. Resuscitation. 2020;148:161-72.
21. Hirayama Y, Ito Y, Ogawa M, Fukushima Y, Ikeyama T. Quantitative assessment of chest compression techniques on an infant manikin. Pediatr Int. 2022;64:e15118.
22. Hirayama Y, Ito Y, Ogawa M, Fukushima Y, Ikeyama T. Quantitative assessment of chest compression techniques on an infant manikin. Pediatrics International: Official Journal of the Japan Pediatric Society. 2022;64:e15118.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.46503
Eurasian J Emerg Med. 2026;25: 135-40

Diagnostic Value of Serum Asymmetric Dimethyl Arginine and Arginine Derivatives Levels in Distinguishing Potentially Life-Threatening Causes in Patients Presenting with Chest Pain to the Emergency Department

Havva Serin Yiğit[1], Ayşegül Bayır[2], Ali Ünlü[2], Fikret Akyürek[2], Bahadır Taşlıdere[3], Ahmet Uyanık[1], Murat Akıncı[1]

[1]Konya Meram State Hospital, Clinic of Emergency Medicine, Konya, Türkiye
[2]Selçuk University Faculty of Medicine, Department of Emergency Medicine and Department of Biochemistry, Konya, Türkiye
[3]Bezmialem Vakıf University Faculty of Medicine, Department of Emergency Medicine, İstanbul, Türkiye

Abstract

Aim: Chest pain a common complaint in the emergency department, often associated with serious cardiovascular conditions such as ischemic heart disease, pulmonary embolism, and aortic aneurysm. Rapid and accurate diagnosis is crucial. Asymmetric dimethylarginine (ADMA), an inhibitor of nitric oxide synthase, contributes to the pathogenesis of cardiovascular disease. Measuring serum ADMA and arginine levels may aid in distinguishing life-threatening conditions and in improving clinical decision-making. This study explores the diagnostic value of ADMA and arginine derivatives in patients presenting with chest pain, aiming to enhance early and effective medical intervention

Materials and Methods: Blood samples were collected from each participant and analyzed for serum ADMA and arginine levels using liquid chromatography-mass spectrometry. Patients were classified based on potentially life-threatening versus benign causes of chest pain. Serum ADMA and arginine levels were compared between the patient and control groups.

Results: A total of 219 participants were included in the study. For biochemical measurements, no significant differences were observed in levels of ADMA, symmetric dimethyl arginine, and NG-monomethyl-l-arginine, whereas arginine and the ratios differed significantly between groups.

Conclusion: The findings offer new approaches to improve early diagnosis and treatment processes. It is anticipated that the results will determine whether the levels of ADMA and arginine can serve as clinically useful adjuncts in the management of chest pain. This could enable the rapid and accurate differentiation of potentially fatal causes, ultimately improving patient management and outcomes.

Keywords: Arginine, ADMA, cardiovascular disease, biomarkers, chest pain

Introduction

Chest pain is one of the most common complaints encountered in emergency departments. It can be a sign of potentially life-threatening conditions, such as myocardial infarction, aortic aneurysm, and pulmonary embolism (PE). Consequently, the rapid and accurate assessment of patients presenting with chest pain is lifesaving. However, a diverse range of potential causes complicates the diagnostic process. In recent years, biomarkers have made significant contributions to clinical practice, particularly the serum markers asymmetric dimethylarginine (ADMA) and arginine. ADMA is an endogenous inhibitor of nitric oxide synthase, and elevated ADMA levels are associated with vascular dysfunction and atherosclerosis (1). Arginine plays a pivotal role in nitric oxide production and supports vascular function. Both molecules may possess substantial diagnostic

Corresponding Author: Havva Serin Yiğit MD, Konya Meram State Hospital, Clinic of Emergency Medicine, Konya, Türkiye
E-mail: hvvsrnygt@gmail.com **ORCID ID:** orcid.org/0009-0009-1196-2057

Cite this article as: Serin Yiğit H, Bayır A, Ünlü A, Akyürek F, Taşlıdere B, Uyanık A, et al. Diagnostic value of serum asymmetric dimethyl arginine and arginine derivatives levels in distinguishing potentially life-threatening causes in patients presenting with chest pain to the emergency department. Eurasian J Emerg Med. 2026;25: 135-40.

Received: 25.03.2025
Accepted: 26.12.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

value in distinguishing cardiovascular events. High levels of ADMA may aid the early diagnosis of cardiovascular diseases, while examining arginine levels can provide insights into vascular function by reflecting nitric oxide levels (2).

This study aims to evaluate the diagnostic value of serum ADMA and arginine levels for differentiating potentially life-threatening causes of chest pain in patients presenting to the emergency department. Studies indicate that ADMA and arginine levels are vital for the diagnosis and management of cardiovascular events. For instance, the work of Zhou et al. (1) highlights the relationship between ADMA and coronary artery disease, demonstrating that elevated ADMA levels are associated with an increased risk of myocardial infarction (1). Numerous studies underscore the diagnostic significance of ADMA and arginine in identifying cardiovascular diseases (3,4).

In this context, evaluating serum ADMA and arginine levels alongside traditional biomarkers may improve the accuracy and effectiveness of diagnostic processes in the emergency department. The findings of this study are intended to contribute significantly to clinical diagnostic algorithms and decision-making processes for emergency medical interventions.

Materials and Methods

This prospective study was conducted between June and November 2016 with approval from the Selçuk University Faculty of Medicine Ethics Committee (decision number: 2016/147 date: 18.05.2016). The study included patients aged 18 years and older who presented with chest pain. Informed consent forms were obtained from all participants. Healthy volunteers without complaints who agreed to participate in the study and showed no pathological findings on physical examination were included as the control group. Patients who met one or more of the following exclusion criteria: chest pain lasting longer than 48 hours; pregnancy or breastfeeding; malignancy; trauma; chronic diseases; or thoracic or esophageal pathologies were excluded from the study. Demographic data, complaints, vital signs, and physical examination findings were recorded. Blood samples were collected from each participant and analyzed for serum ADMA and arginine levels using liquid chromatography-mass spectrometry (LC-MS). Patients were classified according to whether the cause of chest pain was potentially life-threatening (e.g., acute coronary syndrome (ACS), PE, aortic dissection) or benign. Serum ADMA and arginine levels were compared between the patient and control groups. Additionally, medication use, family history, electrocardiogram (ECG) findings, emergency department outcomes, and patients' final status were evaluated.

Quantitative Measurement of ADMA and Other Arginine Derivatives

The normal range for ADMA in adults is accepted to be 0.5-1.2 µmol/L. Serum samples were collected in gel tubes, centrifuged, and stored at -80°C. Plasma samples for measurement of ADMA, symmetric dimethyl arginine (SDMA), NG-monomethyl-l-arginine (L-NMMA), arginine, and citrulline were analyzed at the Selçuk University Faculty of Medicine Biochemistry Laboratory using the AB SCIEX API 3200 LC-MS/MS system. The analysis was conducted in positive-ion mode using Turbo Ion Spray electrospray ionization source on a Phenomenex Luna 50 × 4.6 mm, 5 µm C18 HPLC column. A gradient was established using two mobile phases: water containing 0.1% formic acid (pump A) and methanol containing 0.1% formic acid (pump B). Butanol containing 5% acetyl chloride was used for derivatization. Each sample was processed using prepared reagents, and internal standards were used as necessary (5).

Statistical Analysis

Data analysis was conducted using the SPSS 20.0 statistical software (IBM Inc., Chicago, IL) on the dataset compiled in Microsoft Excel. Descriptive statistics for all variables were calculated. The normality of the data was evaluated using the Kolmogorov-Smirnov test; most of the data did not conform to a normal distribution. Therefore, the Mann-Whitney U test was used to assess differences between two independent groups, and the Kruskal-Wallis test was used for comparisons among multiple groups. Pairwise comparisons were analyzed using the Wilcoxon signed-rank test.

Results

Of the 219 participants included in the study, 47.5% were in the patient group presenting with chest pain and 52.5% were in the control group. Of the participants, 58.4% were male. The cases were classified as ACS, non-specific chest pain (NSGA), PE, and other diagnoses, with NSGA (42.3%) and ACS (35.6%) being the most frequent. The demographic analysis showed similar average ages; the control group had a greater average height, whereas the case group's body mass index (BMI) was 2 kg/m² higher than that of the control group.

Regarding biochemical measurements, no significant differences were found among the levels of ADMA, SDMA, and L-NMMA, whereas arginine ($p<0.001$) and the ratios (arginine/ADMA, $p<0.001$; arginine/total methyl arginine, $p=0.001$) showed significant differences between the groups. The arginine concentration in the patient group was approximately 111 µmol/L, whereas in the control group it was 131 µmol/L. The arginine/ADMA ratio was 403 and 131 µmol/L in the patient

and control groups, respectively (Table 1). All participants in the control group presented in autumn, whereas participants in the patient group presented similarly in summer and autumn. No significant differences were found in the levels of ADMA, SDMA, and L-NMMA. The rates of non-ST elevation myocardial infarction (NSTEMI) and STEMI were notable, and the similarity in pain presentation was 34%.

Cardiovascular risk factors were examined in the patient group: a history of angiography and smoking were present in 40.4%, hypertension in 42.3%, and a family history in 39.4%. Among the patients, 42.2% were discharged after examination, whereas angiography was performed in 33.7%. Among the ECG findings, ST depression was the most commonly observed (15.5%). Presentations occurred during daytime hours in 47.1% of cases. In the initial diagnosis grouping, NSGA (42.3%) and ACS (35.6%) rates were notable, while the PE rate was 8.7%. Troponin and creatine kinase-MB (CK-MB) levels were higher in the NSGA group than in the PE group. Among the arginine derivatives, ADMA levels were lowest in PE and highest in ACS. No mortality was observed in the emergency department; however, three patients died during hospitalization.

In patients with PE, L-NMMA levels were elevated (p=0.013). Simultaneously, SDMA and total methyl arginine increased in cases of cardiomegaly, whereas arginine and the arginine-to-total methyl arginine ratio decreased in cases of atelectasis. In fatal cases, D-dimer, ADMA, SDMA, and L-NMMA levels were elevated; aspartate aminotransferase was elevated in NSGA. D-dimer was significantly elevated in patients with STEMI. Troponin levels did not differ significantly among the groups, with measurements of 100 ng/mL in STEMI, 120 ng/mL in NSTEMI, and 165 ng/mL in unstable angina pectoris (USAP). Four hours later, decreases were observed in STEMI and USAP, while increases were noted in NSTEMI and other groups. CK-MB levels were elevated in NSTEMI and other groups. D-dimer levels were significantly elevated in patients with STEMI. The arginine/ADMA (p=0.017) and arginine/total methyl arginine (p=0.037) ratios showed significant differences among the STEMI, NSTEMI, and USAP groups. Arginine/ADMA was 642 µmol/L in STEMI, 318 µmol/L in NSTEMI, and 176 µmol/L in USAP. Arginine/total methyl arginine was 313 µmol/L in STEMI, 187 µmol/L in NSTEMI, and 104 µmol/L in USAP. ADMA levels were elevated in NSTEMI and USAP but reduced in STEMI. SDMA concentrations were 314 µmol/L in NSTEMI and 172 µmol/L in STEMI. L-NMMA was elevated in USAP to 0.045 µmol/L. Citrulline concentration peaked at 23.78 µmol/L in USAP. Total methylarginine values were similar in NSTEMI and USAP (both 0.700 µmol/L) (Table 2).

In patients with a history of angiography, levels of troponin, CK-MB, and D-dimer were lower, whereas levels of ADMA and total methyl arginine were higher. In patients with typical pain, CK-MB levels were significantly higher. SDMA levels were substantially higher in those with a history of heart failure (p=0.040). Smokers had lower citrulline levels (p=0.038), whereas physically active individuals had lower SDMA and total methylarginine levels (p=0.045 and p=0.009, respectively). In those with a history of hypertension, troponin and citrulline levels were elevated (p=0.006 and p=0.009, respectively). As the duration of the stay in the intensive care unit increased, SDMA levels rose while arginine levels decreased. A positive correlation was found between ADMA and total methyl arginine, whereas an inverse correlation was observed between ADMA and the arginine/ADMA ratio. The ROC analysis of the arginine-to-ADMA ratio demonstrated high diagnostic accuracy (area under the curve: 0.867) with a cutoff value of 218.15 µmol/L. Alcohol use was more common in the group with fatal cardiovascular disease (p=0.019). Most of the fatal cases were male and presented during daytime hours (Table 3, Figure 1).

Table 1. Arginine derivatives values in patients

Parameter	Patients (n=104)	Control (n=115)	p value
ADMA (µmol/L)	0.315±0.133	0.327±0.136	0.515
SDMA (µmol/L)	0.229±0.182	0.189±0.080	0.184
L-NMMA (µmol/L)	0.038±0.034	0.034±0.014	0.567
Arginine (µmol/L)	111.34±63.33	131.52±44.90	<0.001*
Citrulline (µmol/L)	18.76±15.37	19.44±11.33	0.103
Total methyl arginine (µmol/L)	0.583±0.261	0.551±0.182	0.556
Arginine/ADMA	403.19±318.09	131.75±44.91	<0.001*
Arginine/total methyl arginine	218.36±160.09	151.0±48.3	0.001*

*: p<0.001, ADMA: Asymmetric dimethylarginine, SDMA: Symmetric dimethyl arginine, L-NMMA: NG-monomethyl-l-arginine

Table 2. Biochemical findings of cardiac markers according to detailed diagnostic categories					
Parameter	**STEMI (n=11)**	**NSTEMI (n=22)**	**USAP (n=4)**	**Other (n=67)**	**p value**
Troponin (ng/L)	98.38±250.5	121.32±298	164.9±186.2	130.4±330	0.703
CK-MB (ng/mL)	3.44±2.56	5.60±12.5	3.05±1.8	9.74±26.3	0.886
D-dimer (ng/mL)	2520±353	1963±2439	-	-	0.946
INR	1.31	1.27	-	-	0.984
ADMA (µmol/L)	0.254±0.12	0.328±0.14	0.381±0.12	0.315±0.13	0.384
SDMA (µmol/L)	0.172±0.04	0.314±0.36	0.273±0.11	0.208±0.10	0.150
L-NMMA (µmol/L)	0.034±0.01	0.041±0.02	0.045±0.01	0.038±0.04	0.942
Arginine (µmol/L)	139.6±104.7	102.5±65.7	68.17±40.8	112.5±54.8	0.151
Citrulline (µmol/L)	19.63±11.4	18.96±23.1	23.78±24.3	18.27±13.9	0.941
Total methyl arginine (µmol/L)	0.460±0.16	0.692±0.20	0.700±0.02	0.562±0.21	0.122
Arginine/ADMA	642.0±508.7[a,b]	318.7±135.3[b]	176.8±107.6[a]	407.1±308.4	0.017*
Arginine/total methyl arginine	313.3±210.3[a,b]	187.8±88.9[b]	104.1±75.3[a]	227.1±163.9	0.037*

Significant differences marked with asterisks. "a" and "b" represent groups with significant differences (comparison indicated in the table
CK-MB: Creatine kinase-MB INR: International normalized ratio, ADMA: Asymmetric dimethylarginine, SDMA: Symmetric dimethyl arginine, L-NMMA: NG-monomethyl-l-arginine, NSTEM: non-ST elevation myocardial infarction, STEM: ST elevation myocardial infarction

Table 3. Differential diagnostic information determined for the arginine/ADMA ratio			
Arginin/ADMA	**Patients**	**Control**	**Total**
>218.16	78	5	83
<218.15	26	110	136
Total	104	115	219
Sensitivity	75.0%	False negative	25.0%
Specificity	95.65%	False positive	4.34%
Positive predictive value	93.97%	Positive probability	1725%
Negative predictive value	80.88%	Negative probability	26.13%
Accuracy	85.84%		
ADMA: Asymmetric dimethylarginine			

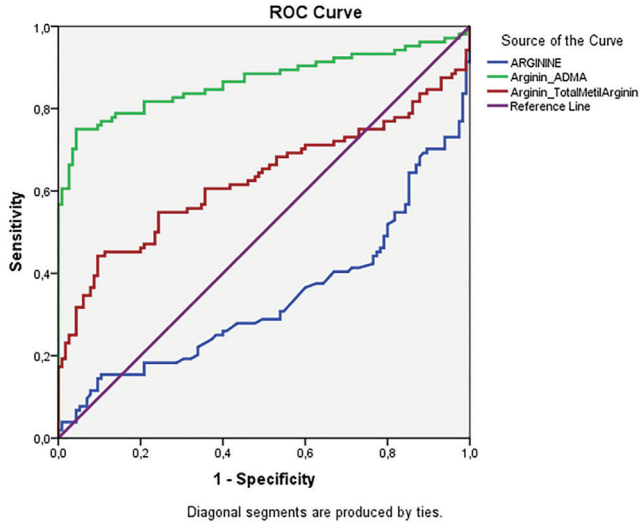


Figure 1. Results of ROC analysis for arginine and its ratios
ADMA: Asymmetric dimethylarginine

Discussion

This study investigated the relationship between arginine derivatives (ADMA, SDMA, L-NMMA) and cardiovascular diseases. Elevated plasma ADMA levels are an important risk factor for atherosclerosis and cardiovascular events. The finding of lower citrulline levels in smokers highlights the effects of smoking on metabolism. A positive correlation has been observed between age and ADMA levels, indicating that ADMA levels increase with age. The study by Cavusoglu et al. (6) indicates that high levels of ADMA and SDMA can serve as cardiovascular risk factors and are associated with potentially fatal cardiovascular events. These findings are consistent with previous studies (7,8).

Previous research has examined the effects of arginine derivatives on cardiovascular disease. Borgeraas et al. (9) found that ADMA levels were higher in the low BMI group, indicating that arginine derivatives are associated with increased cardiovascular risk. In another study conducted by Borgeraas et al. (9), male gender and age were associated with ADMA levels in patients grouped by trans fatty acids. These findings align with the results of this study and strengthen the relationship between arginine and ADMA levels and cardiovascular diseases.

Previous studies have shown that ADMA and SDMA could serve as cardiovascular risk factors (10). Souza-Costa et al. (11) examined the effects of L-NMMA on pulmonary hypertension and provided evidence of a relationship between arginine derivatives and cardiovascular conditions. Böger et al. (3) emphasized the significance of the relationship between ADMA and cardiovascular diseases. In our study, elevated levels of ADMA were observed in patients with chest pain admitted to the emergency department. Cavusoglu et al. (6) also examined the relationship between ADMA and BMI, and found a significant association between obesity and elevated ADMA levels. Our study similarly demonstrates that ADMA is associated with BMI, hypertension, and other cardiovascular risk factors, consistent with previous studies.

Das et al. (12) examined the effects of hypertension and diabetes on ADMA levels and observed significantly elevated levels in patients with hypertension. Our study also found increases in arginine and SDMA levels in patients with a history of hypertension. Zoccali et al. (13) pointed out that inflammation increases ADMA levels; our study also demonstrated an association between inflammatory conditions and ADMA. Kumar et al. (14) examined the relationship between pulmonary hypertension and L-NMMA, indicating that L-NMMA could be used for clinical diagnosis. Lippi et al. (15) investigated the diagnostic value of arginine derivatives in ACS and noted the elevation of the arginine/ADMA ratio. These findings suggest that arginine derivatives may be important in clinical diagnosis.

A major limitation of this study is the small sample size in certain diagnostic subgroups, particularly the USAP group (n=4). This small subgroup size reduces the statistical power and limits the reliability of subgroup comparisons. Therefore, the findings related to USAP should be interpreted with caution and considered hypothesis-generating rather than definitive.

Study Limitations

The study's generalizability is limited by the small sample size and heterogeneous group composition. Larger population-based, long-term studies are needed to understand the value of arginine derivatives better. To reduce confounding, patients with conditions known to elevate ADMA independently of cardiovascular disease—such as renal failure, infection/sepsis, systemic inflammatory or autoimmune disorders, malignancy, and chronic liver disease—were excluded. This ensured that observed changes in ADMA and arginine metabolism primarily reflected cardiovascular pathology.

Conclusion

Elevated ADMA levels and an increased arginine/ADMA ratio have been identified as important biomarkers in the differential diagnosis of cardiovascular events and fatal conditions. Findings emphasize the connection between ADMA and factors such as inflammation, hypertension, and obesity, suggesting potential contributions to early diagnosis and treatment. The use of the arginine/ADMA ratio in diagnostic processes is recommended. This study highlights that these biomarkers could be critical tools in emergency management and lays the groundwork for future research.

Ethics

Ethics Committee Approval: The study was conducted between June and November 2016 and was approved by the Selçuk University Faculty of Medicine Ethics Committee (decision no: 2016/147, date: 18.05.2016).

Informed Consent: Informed consent forms were obtained from all participants.

Footnotes

Authorship Contributions

Concept: HSY., Design: HSY, AB., Data Collection or Processing: HSY, AB, AU., Analysis or Interpretation: HSY, AB, FA., Literature Search: HSY, AB, AU, MA, BT., Writing: HSY.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Zhou Z, Jing H, Tao Z, Choi H, Aboulfatova K, Moake J, et al. Effects of naturally occurring mutations in CUB-1 domain on synthesis, stability, and activity of ADAMTS-13. Thromb Res. 2009;124:323-7.
2. Cotter G, Kaluski E, Blatt A, Milovanov O, Moshkovitz Y, Zaidenstein R, et al. L-NMMA (a nitric oxide synthase inhibitor) is effective in the treatment of cardiogenic shock. Circulation. 2000;101:1358-61.
3. Böger RH. Association of asymmetric dimethylarginine and endothelial dysfunction. Clin Chem Lab Med. 2003;41:1467-72.
4. Pettersson A, Hedner T, Milsom I. Increased circulating concentrations of asymmetric dimethyl arginine (ADMA), an endogenous inhibitor of nitric oxide synthesis, in preeclampsia. Acta Obstet Gynecol Scand. 1998;77:808-13.
5. Schulze F, Wesemann R, Schwedhelm E, Sydow K, Albsmeier J, Cooke JP, et al. Determination of asymmetric dimethylarginine (ADMA) using a novel ELISA assay. Clin Chem Lab Med. 2004;42:1377-83.
6. Cavusoglu E, Ruwende C, Chopra V, Poludasu S, Yanamadala S, Frishman WH, et al. Relation of baseline plasma ADMA levels to cardiovascular morbidity and mortality at two years in men with diabetes mellitus referred for coronary angiography. Atherosclerosis. 2010;210:226-31.

7. Kielstein JT, Salpeter SR, Bode-Boeger SM, Cooke JP, Fliser D. Symmetric dimethylarginine (SDMA) as endogenous marker of renal function--a meta-analysis. Nephrol Dial Transplant. 2006;21:2446-51.
8. Valkonen M, Penttilä M, Saloheimo M. Effects of inactivation and constitutive expression of the unfolded- protein response pathway on protein production in the yeast saccharomyces cerevisiae. Appl Environ Microbiol. 2003;69:2065-72.
9. Borgeraas H, Hertel JK, Svingen GF, Pedersen ER, Seifert R, Nygård O, et al. Association between body mass index, asymmetric dimethylarginine and risk of cardiovascular events and mortality in norwegian patients with suspected stable angina pectoris. PLoS One. 2016;11:e0152029.
10. Hov GG, Aasarød KI, Sagen E, Åsberg A. Arginine, dimethylated arginine and homoarginine in relation to cardiovascular risk in patients with moderate chronic kidney disease. Clin Biochem. 2015;48:646-51.
11. Souza-Costa DC, Zerbini T, Metzger IF, Rocha JB, Gerlach RF, Tanus-Santos JE. l-Arginine attenuates acute pulmonary embolism-induced oxidative stress and pulmonary hypertension. Nitric Oxide. 2005;12:9-14.
12. Das UN, Repossi G, Dain A, Eynard AR. L-arginine, NO and asymmetrical dimethylarginine in hypertension and type 2 diabetes. Front Biosci (Landmark Ed). 2011;16:13-20.
13. Zoccali C, Maas R, Cutrupi S, Pizzini P, Finocchiaro P, Cambareri F, et al. Asymmetric dimethyl-arginine (ADMA) response to inflammation in acute infections. Nephrol Dial Transplant. 2007;22:801-6.
14. Kumar R, Gupta N, Jain P. L-NMMA levels and their association with pulmonary hypertension. Respir Med. 2023;121:95-101.
15. Lippi G, Montagnana M, Salvagno GL, Guidi GC. Potential value for new diagnostic markers in the early recognition of acute coronary syndromes. CJEM. 2006;8:27-31.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.33600
Eurasian J Emerg Med. 2026;25: 141-6

Clinical Characteristics of Pulmonary Contusion in Trauma Patients and Determinants of Mortality: A Retrospective Analysis

Abdul Samet Şahin[1], Serkan Özden[2], Emre Dilaver[1], Buket Kaytaz Alkaş[2], İrem Dilaver[3], Sinan Paslı[1]

[1]Karadeniz Technical University Faculty of Medicine, Department of Emergency Medicine, Trabzon, Türkiye
[2]Karadeniz Technical University Faculty of Medicine, Department of Thoracic Surgery, Trabzon, Türkiye
[3]Karadeniz Technical University Faculty of Medicine, Department of Public Health Trabzon, Türkiye

Abstract

Aim: Pulmonary contusion is a leading cause of morbidity and mortality in trauma patients. This study retrospectively evaluated the clinical characteristics of patients with pulmonary contusion who presented to the emergency department, as well as the factors associated with mortality.

Materials and Methods: This retrospective, cross-sectional study included patients aged 18 years and older who were diagnosed with pulmonary contusion by computed tomography and who presented to the emergency department of a tertiary university hospital between 2019 and 2024. The analysis included demographic characteristics, trauma mechanism, associated injuries, including by body region, laboratory parameters, complications, treatment approaches, and clinical outcomes.

Results: A total of 350 patients were evaluated; mean age was 47.8±18.6 years, and 79.4% were male. The most common trauma mechanism was blunt trauma. Pulmonary contusion was most frequently observed in the lower lobes. The most prevalent associated injuries were rib fractures, pneumothorax, and hemothorax. The 30-day total mortality rate was 8%. Among non-survivors, the following were significantly higher: age, chest trauma scores, complication rates, lactate levels, and pCO_2 levels.

Conclusion: In trauma patients, pulmonary contusion most frequently occurs in the lower lobes and commonly accompanies chest wall and pleural injuries. Advanced age, a higher chest trauma score, the development of complications, and elevated lactate and pCO_2 levels are risk factors for mortality. Early evaluation of these parameters may facilitate prompt identification of high-risk cases.

Keywords: Pulmonary contusion, trauma, thoracic injury, mortality, emergency department

Introduction

Trauma is a significant cause of morbidity and mortality among young and working-age individuals (1). Blunt thoracic trauma, in particular, accounts for a significant number of emergency department visits and hospital admissions. It can lead to severe intrathoracic injuries, including pulmonary contusion, pneumothorax, hemothorax, and rib fractures (2,3). The most common parenchymal injury in thoracic trauma is pulmonary contusion, which leads to respiratory failure through pathophysiological processes such as alveolar hemorrhage, edema, ventilation–perfusion mismatch, and intrapulmonary shunting. This condition not only causes local complications but may also trigger a systemic inflammatory response, thereby adversely affecting patient prognosis (4).

Pulmonary contusions are reported in 17-75% of blunt chest trauma cases and are considered an independent risk factor for mortality (5,6). Pathologies accompanying pulmonary contusion, such as rib fractures, pneumothorax, and hemothorax, increase morbidity and the need for invasive interventions (4). Additionally, the development of complications such as pneumonia and acute respiratory distress syndrome (ARDS) is one of the main factors that worsen the clinical course and increase mortality (7). Nevertheless, the clinical course of pulmonary contusion

Corresponding Author: Abdul Samet Şahin MD, Karadeniz Technical University Faculty of Medicine, Department of Emergency Medicine, Trabzon, Türkiye
E-mail: abdulsahin61@gmail.com **ORCID ID:** orcid.org/0000-0001-9512-8741

Cite this article as: Şahin AS, Özden S, Dilaver E, Kaytaz Alkaş B, Dilaver İ, et al. Clinical characteristics of pulmonary contusion in trauma patients and determinants of mortality: a retrospective analysis. Eurasian J Emerg Med. 2026;25: 141-6.

Received: 25.09.2025
Accepted: 27.12.2025
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

is unpredictable, and varying outcomes are observed among patients (8).

Therefore, clarifying the clinical characteristics of pulmonary contusion and the factors influencing prognosis is important for enabling early risk stratification. This study aimed to retrospectively evaluate the demographic characteristics, trauma mechanisms, associated injuries, laboratory parameters, treatment approaches, and clinical outcomes of patients with pulmonary contusion presenting to the emergency department, and to identify determinants associated with mortality.

Materials and Methods

Study Design

This study was designed as a retrospective, cross-sectional analysis. Patients who presented to the emergency department of a tertiary university hospital between January 1, 2019, and December 31, 2024, and who were referred to the thoracic surgery clinic from the emergency department were included if pulmonary contusion was diagnosed by computed tomography (CT) (i.e., CT reports confirming pulmonary contusion). The study group consisted of all cases recorded within the specified time frame that met the inclusion criteria. For the analysis, only patients aged 18 years and older with complete and verifiable medical records were included. Cases with incomplete or unverifiable records, patients younger than 18 years, and patients referred to another healthcare facility for whom follow-up data were unavailable were excluded from the study.

The Institutional Review Board of Karadeniz University, Faculty of Medicine, approved the study (decision number: 2025/125, date: 14.05.2025). The study had a retrospective design, and patient data were obtained solely from medical records. Patient names and identifying information were not used; all data were anonymized before analysis. The study was conducted in accordance with the principles of the Declaration of Helsinki, and patient confidentiality was maintained.

As this was a retrospective study, informed consent was waived. Patient records were anonymized, and no personal identifiers were used.

Data Collection Process

Eligible patients were identified through the hospital's information management system and Picture Archiving and Communication System. Patient lists were generated using the ICD-10 code S20.2 (Thorax Contusion). Additionally, patients referred from the emergency department to the thoracic surgery clinic were included in the study group. The collected variables included:

- Demographic data: age and sex.
- Trauma characteristics: mechanism of trauma (blunt, penetrating, fall from height, other).
- Associated injuries: rib fracture, sternal fracture, clavicle fracture, pneumothorax, hemothorax, contusion, subcutaneous emphysema, pneumomediastinum, diaphragmatic injury, pulmonary laceration, and cardiac contusion.
- Laboratory parameters included blood gases (pO_2, pCO_2), hemoglobin, lactate, and troponin.
- Treatment process: conservative management, tube thoracostomy, thoracotomy, and need for surgical intervention.
- Clinical outcomes include discharge, ward admission, intensive care admission, hospital length of stay, development of complications, and mortality.
- Chest injury score: Chest trauma score (CTS)(9).
- Contusion localization: right upper, right lower, right middle, left upper, and left lower lung zones.

All data were recorded in an electronic spreadsheet, verified using a double-check method, and prepared for analysis.

Statistical Analysis

Data analysis was performed using Jamovi software (version 2.6.24.0). The distribution of continuous variables was assessed using the Kolmogorov-Smirnov test. Continuous variables with a normal distribution were expressed as the mean ± standard deviation, while those without a normal distribution were expressed as the median (minimum-maximum). Categorical variables were presented as numbers and percentages.

For comparisons between groups, the Independent Samples t-test was used for continuous variables with normal distribution, while the Mann-Whitney U test was applied for those without normal distribution. The chi-square test or Fisher's exact test, as appropriate, was used to analyze categorical variables. A p value <0.05 was considered statistically significant in all analyses.

Results

A total of 350 patients were included in the study. The mean age was 47.8±18.6 years, and 79.4% of the patients were male. The majority of patients sustained blunt trauma (96.3%, n=337), followed by penetrating trauma (3.7%, n=13). The most common associated injuries were rib fractures (68.9%), pneumothorax (44.3%), and hemothorax (28.6%). Severe intrathoracic injuries, such as diaphragmatic rupture or esophageal injury, were infrequent (less than 1%). The most frequently injured sites

were the right lower lobe (57.7%) and the left lower lobe (54.3%). The overall 30-day mortality rate was 8% (n=28). A subgroup analysis revealed that mortality was exclusively observed in the polytrauma group (10.4%, n=28), whereas no deaths occurred in the isolated pulmonary contusion group (0%, n=0). Among the 350 patients with pulmonary contusion, the majority had associated extra-thoracic injuries. The most frequently affected body regions were the extremities (39.1%, n=137) and the spine (34.0%, n=119). Other associated injuries included traumatic brain injuries n=(23.4%, n =82), maxillofacial injuries (18.9%, n=66), abdominal injuries (16.0%, n=56), and pelvic fractures (10.9%, n=38). Less common injuries were vascular injuries (1.4%, n=5) and cardiac injuries (0.6%, n=2). The demographic and clinical characteristics of the patients are presented in Table 1.

The mean age of deceased patients was significantly higher than that of survivors: 57±19.3 years versus 47±18.2 years (p=0.006). There was no significant difference in sex distribution between the groups (p=0.712). The proportion of patients with a CTS

Table 1. Demographic ınformation and clinical characteristics

	Age (mean ± SD)	n (%)
Gender		
Male	45.7±18	278 (79.4%)
Female	55.7±18.7	72 (20.6%)
Trauma characteristics		**n (%)**
Blunt trauma		337 (96.3%)
Penetrating trauma		13 (3.7%)
Patient group		
Isolated thorax trauma		80 (22.9%)
Polytrauma		270 (77.1%)
Associated injuries		
Rib fracture		241 (68.9%)
Pneumothorax		155 (44.3%)
Hemothorax		100 (28.6%)
Subcutaneous emphysema		57 (16.3%)
Clavicle fracture		30 (8.6%)
Pneumomediastinum		18 (5.1%)
Pulmonary laceration		18 (5.1%)
Sternal fracture		17 (4.9%)
Diaphragmatic injury		1 (0.3%)
Contusion location		
Right lower lobe		202 (57.7%)
Left lower lobe		190 (54.3%)
Right upper lobe		113 (32.3%)
Left upper lobe		113 (32.3%)
Right middle lobe		77 (22.0%)
Associated ınjuries by body region		
Extremity		137 (39.1%)
Spinal		119 (34.0%)
Brain		82 (23.4%)
Maxillofacial		66 (18.9%)
Abdominal		56 (16.0%)
Pelvis		38 (10.9%)
Vascular		5 (1.4%)
Cardiac		2 (0.6%)

Table 1. Continued

	n (%)
Complications	
Total number of patients with complications	98 (28.0%)
Pneumonia	22 (6.3%)
Pneumothorax	15 (4.3%)
Hemothorax	16 (4.6%)
Atelectasis	5 (1.4%)
ARDS	4 (1.1%)
Hemoptysis	1 (0.3%)
Other	58 (16.6%)
Treatment methods	
Observation/follow-up	268 (76.8%)
Tube thoracostomy	77 (22.1%)
Surgery	4 (1.1%)
Hospitalization and outcomes	
Number of hospitalized patients	288 (82.3%)
Discharged from ED	54 (15.4%)
Ward admission	170 (48.6%)
ICU admission	116 (33.1%)
Death in ED	10 (2.9%)
30-day mortality	
Isolated thorax trauma	0 (0%)
Total polytrauma	28 (8.0%)
	Median (minimum-maximum)
Average length of stay (days)	5 (0-67)

ED: Emergency department, ICU: Intensive care unit, SD: Standard deviation, ARDS: Acute respiratory distress syndrome

greater than five was 92.9% in the mortality group and 56.2% among survivors, a statistically significant difference (p<0.001). Complications during clinical follow-up also occurred more frequently in the mortality group (89.3%) than in the survivor group (22.7%), with a statistically significant difference (p<0.001).

We then compared the patients who survived (n=322) with those who died (n=28) to identify factors associated with mortality. When examining treatment methods, tube thoracostomy was required significantly more often in the mortality group than among survivors (p=0.002). Regarding laboratory parameters, the median lactate level was 28 mg/dL in patients who died, compared with 15 mg/dL in survivors (p=0.001). Similarly, pCO_2 levels were higher in the mortality group (p<0.001). In contrast, troponin, pO_2, and hemoglobin values were not significantly associated with mortality (p>0.05). A comparison of variables associated with mortality is summarized in Table 2.

Discussion

This study retrospectively evaluated the clinical characteristics of patients with pulmonary contusion who presented to the emergency department and the factors associated with mortality. Our findings suggest that advanced age, CTS, and the development of complications are key determinants of mortality. These results demonstrate both similarities and differences relative to previous studies of factors influencing the prognosis of pulmonary contusion.

In this study, which evaluated 350 patients, the majority of patients were male, with a mean age of 47.8 years. Similarly, in a 10-year retrospective study conducted by Demirhan et al., the proportion of male patients was 85%, and the mean age was 36.2 years (10). This finding is consistent with the literature reporting that trauma occurs among young and middle-aged males (3,11).

Examination of trauma mechanisms showed that nearly all cases were caused by blunt-force trauma or falls, whereas penetrating injuries occurred at very low rates. Previous studies have reported that most thoracic injuries result from blunt force, with traffic accidents and falls from heights being the most common causes (2,11-13). In our study, falls also emerged as the most significant mechanism associated with blunt trauma.

Pulmonary contusions were most frequently localized in the lower lobes: the right lower lobe (57.7%) and the left lower lobe (54.3%). This may be explained by the larger anatomical volume of the lower lobes and their greater exposure to trauma (14). The literature also reports that the lower lobes are more susceptible to contusion (13,15).

The most common associated injuries were rib fractures, pneumothorax, and hemothorax. These findings demonstrate that pulmonary contusion is frequently accompanied by chest wall and pleural pathologies. According to Tyburski et al. (16), rib fractures were the most frequent associated injury, occurring in 67% of pulmonary contusion cases. In a study by Dumanlı evaluating pulmonary contusions, pneumothorax was detected in 51.9% of cases and hemothorax was detected in 45.4% (2). Our findings are consistent with previous research.

When complications were evaluated, 28% of cases exhibited complications; the most prominent were respiratory problems, including pneumonia (6.3%), pneumothorax (4.3%), hemothorax (4.6%), and ARDS (1.1). Pulmonary contusion has also been reported in the literature as an important risk factor, particularly for the development of pneumonia and ARDS (5,7,11,17). In the study by Lee et al. (18), 38% of patients with contusions developed pneumonia, while 6.8% developed ARDS. Our study, observed lower rates, possibly owing to effective patient monitoring, early diagnosis and intervention, and the widespread use of

Table 2. Comparison of clinical and laboratory findings according to mortality status

Variable	Mortality (-) (n = 322)	Mortality (+) (n = 28)	p-value
Age (years, mean ± SD)	47.0±18.2	57.1±19.3	0.006
Gender (M/F)	256/66	22/6	0.712
Complication (%)	22.7	89.3	<0.001
CTS score >5 (%)	56.2	92.9	<0.001
Treatment: tube thoracostomy (%)	19.6	53.6	0.002
Troponin (ng/L, median)	19.3 (3.1-4761.0)	18.8 (3.5-216.1)	0.632
pCO_2 (mmHg, mean ± SD)	41.0±9.5	47.0±11.2	<0.001
pO_2 (mmHg, mean ± SD)	69.2±58.0	65.8±63.4	0.253
HGB (g/dL, mean ± SD)	13.7±2.0	13.3±2.1	0.053
Lactate (mg/dL, median)	14.0 (0.7-219.0)	28.5 (0.9-81.0)	0.001

SD: Standard deviation, M/F: Male/female, CTS: Chest trauma score, HGB: Hemoglobin

conservative treatment approaches. However, the development of respiratory complications remains an important factor in determining the clinical course. Close monitoring and effective management of these complications are critical to the success of treatment in such patients.

Examination of laboratory findings revealed that lactate levels were significantly higher in the mortality group. However, troponin and hemoglobin values were not associated with mortality. Previous studies have also emphasized the importance of lactate levels as prognostic biomarkers (19-21). The higher pCO_2 values observed in the mortality group demonstrate the impact of ventilatory impairment on prognosis (22). These results show that respiratory failure is a critical factor in determining mortality in patients with pulmonary contusion. Elevated lactate levels, an indicator of tissue hypoperfusion and oxygen deficiency, stand out among laboratory parameters as strong biomarkers of mortality (19). Elevated pCO_2 values reflect impaired alveolar ventilation and demonstrate the impact of respiratory failure on mortality. In contrast, troponin, hemoglobin, and pO_2 levels were not significantly associated with mortality. These results suggest that cardiac biomarkers and oxygenation parameters may play only a secondary role in contusion-related mortality (23). These findings are consistent with those of earlier studies that identified similar parameters as risk factors for mortality (24).

In our study, when treatment approaches were evaluated, the majority of cases were managed conservatively. However, the rate of tube thoracostomy was significantly higher among patients who died. This finding suggests that pleural complications negatively affect prognosis and increase the need for invasive interventions. Most patients with thoracic trauma recover clinically within 5-7 days with conservative management; however, a subset may require invasive treatment (5). In particular, tube thoracostomy is the first-line intervention for managing complications of pulmonary contusion. At the same time, the need for surgical intervention arises in a limited number of cases (5,8). Accordingly, the low rate of surgical treatment in our study is consistent with existing literature. However, a greater frequency of invasive interventions in cases of high CTS with complications underscores the importance of early risk assessment in treatment strategies.

Our study found a mortality rate of 8%, which is notably lower than the 14-40% reported in the literature for pulmonary contusion (5,12,16). Reasons for this difference may include the centre's advanced intensive care facilities, early diagnosis and effective resuscitation strategies, a predominance of younger patients, and a relatively low incidence of major concomitant trauma. Examination factors affecting mortality identified advanced age, the development of complications, higher CTS, and elevated lactate and pCO_2 levels as the strongest determinants of mortality. Reduced physiological reserves, increased prevalence of comorbidities, and diminished recovery capacity can explain the impact of advanced age on mortality (5). Similarly, the presence of complications, particularly pneumonia and ARDS, prolongs the hospital stay, worsens the clinical course, and increases mortality (8). The correlation between a high CTS and mortality underscores the pivotal role of trauma severity and associated injuries in informing prognosis (25). Therefore, early evaluation of age, trauma score, lactate, and pCO_2 levels in patients with pulmonary contusion may facilitate the rapid identification of high-risk cases and contribute to the implementation of treatment strategies aimed at improving prognosis.

Study Limitations

This study has several limitations. First, because of its retrospective design, data were obtained from existing medical records; therefore, the possibility of missing or incomplete information cannot be excluded. Second, since the study was conducted at a single center, the results may not be generalizable to other patient populations. Third, while we utilized the CTS to assess thoracic injury severity, global trauma scores such as the injury severity score or TRISS were not calculated. The impact of multi-trauma was instead evaluated through the analysis of specific associated injuries. Additionally, the clinical effects of trauma severity and associated injuries may not have been evaluated uniformly.

Conclusion

Pulmonary contusion most commonly occurs after blunt trauma or falls and is frequently accompanied by rib fractures, pneumothorax, and hemothorax. In our study, the mortality rate was 8%; advanced age, development of complications, higher CTS, elevated pCO_2, and increased lactate levels were identified as risk factors. These results highlight the critical importance of early clinical and laboratory assessments in the emergency department for identifying high-risk patients and planning appropriate treatment strategies.

Ethics

Ethics Committee Approval: The Institutional Review Board of Karadeniz University, Faculty of Medicine, approved the study (decision number: 2025/125, date: 14.05.2025). The study had a retrospective design, and patient data were obtained solely from medical records. Patient names and identifying information were not used; all data were anonymized before analysis. The study was conducted in accordance with the principles of the Declaration of Helsinki, and patient confidentiality was maintained.

Informed Consent: As this was a retrospective study, informed consent was waived. Patient records were anonymized, and no personal identifiers were used.

Footnotes

Authorship Contributions

Concept: A.S.S., B.K.A., İ.D., Design: A.S.S., E.D., İ.D., S.P., Data Collection or Processing: S.Ö., B.K.A., İ.D., S.P., Analysis or Interpretation: A.S.S., E.D., B.K.A., İ.D., Literature Search: S.Ö., S.P., Writing: A.S.S., S.Ö., E.D., S.P.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Kanake V, Kale K, Mangam S, Bhalavi V. Thorax trauma severity score in patient with chest trauma: study at tertiary-level hospital. Indian J Thorac Cardiovasc Surg. 2022;38:149-56.
2. Dumanli A. Künt toraks travması geçiren hastalarda oluşan akciğer kontüzyonu ile hemotoraks / pnömotoraks arasındaki ilişki. Kocatepe Medical Journal. 2020;21:217-21.
3. Hasbahçeci M, Özpek A, Başak F, Çalişkan M, Ener BK, Alimoğlu O. Künt toraks travmasında mortaliteye etki eden faktörler. Ulusal Travma ve Acil Cerrahi Dergisi. 2013;19:127-32.
4. Arjun P, Najeeb AA, Ramachandran R. Lung contusion. Lung India. 2018;35:263-4.
5. Dogrul BN, Kiliccalan I, Asci ES, Peker SC. Blunt trauma related chest wall and pulmonary injuries: an overview. Chin J Traumatol. 2020;23:125-38.
6. Bader A, Rahman U, Morris M, McCormack JE, Huang EC, Zawin M, et al. Pulmonary contusions in the elderly after blunt trauma: incidence and outcomes. J Surg Res. 2018;230:110-6.
7. Miller PR, Croce MA, Bee TK, Qaisi WG, Smith CP, Collins GL, et al. ARDS after pulmonary contusion: accurate measurement of contusion volume identifies high-risk patients. J Trauma. 2001;51:223-8; discussion 229-30.
8. Rendeki S, Molnár TF. Pulmonary contusion. J Thorac Dis. 2019;11(Suppl 2):S141-S151.
9. Chen J, Jeremitsky E, Philp F, Fry W, Smith RS. A chest trauma scoring system to predict outcomes. Surgery. 2014;156:988-93.
10. Demirhan R, Onan B, Oz K, Halezeroglu S. Comprehensive analysis of 4205 patients with chest trauma: a 10-year experience. Interact Cardiovasc Thorac Surg. 2009;9:450-3. Epub 2009 Jun 18.
11. Yetim TD, Yetim İ, Duru M, Bayaroğulları H. Thoracic trauma: 156 patient assessment. Eurasian Journal of Emergency Medicine. 2011;10:110-3.
12. Battle C, Carter K, Newey L, Giamello JD, Melchio R, Hutchings H. Risk factors that predict mortality in patients with blunt chest wall trauma: an updated systematic review and meta-analysis. Emerg Med J. 2023;40:369-78.
13. Dennis BM, Bellister SA, Guillamondegui OD. Thoracic trauma. Surg Clin North Am. 2017;97:1047-64.
14. Yamada Y, Yamada M, Yokoyama Y, Tanabe A, Matsuoka S, Niijima Y, et al. Differences in lung and lobe volumes between supine and standing positions scanned with conventional and newly developed 320-detector-row upright CT: intra-individual comparison. Respiration. 2020;99:598-605.
15. Ganie FA, Lone H, Lone GN, Wani ML, Singh S, Dar AM, et al. Lung contusion: a clinico-pathological entity with unpredictable clinical course. Bull Emerg Trauma. 2013;1:7-16.
16. Tyburski JG, Collinge JD, Wilson RF, Eachempati SR. Pulmonary contusions: quantifying the lesions on chest X-ray films and the factors affecting prognosis. J Trauma. 1999;46:833-8.
17. Horst K, Simon TP, Pfeifer R, Teuben M, Almahmoud K, Zhi Q, et al. Characterization of blunt chest trauma in a long-term porcine model of severe multiple trauma. Sci Rep. 2016;6:39659.
18. Lee NH, Kim SH, Seo SH, Kim BJ, Lee CS, Kim GH, et al. Prediction of respiratory complications by quantifying lung contusion volume using chest computed tomography in patients with chest trauma. Sci Rep. 2023;13:6387.
19. Çınar E, Usul E, Demirtaş E, Gökçe A. The role of trauma scoring systems and serum lactate level in predicting prognosis in thoracic trauma. Ulus Travma Acil Cerrahi Derg. 2021;27:619-23.
20. Guyette F, Suffoletto B, Castillo JL, Quintero J, Callaway C, Puyana JC. Prehospital serum lactate as a predictor of outcomes in trauma patients: a retrospective observational study. J Trauma. 201;70:782-6.
21. Yazıcı R, Bulut B, Genc M, Akkan Öz M, Hanalioglu D, Kokulu K, et al. Prognostic indicators in patients with isolated thoracic trauma: a retrospective cross-sectional study. Ulus Travma Acil Cerrahi Derg. 2024;30:737-44.
22. Corral Torres E, Hernández-Tejedor A, Millán Estañ P, Valiente Fernández M, Bringas Bollada M, Pérez Díaz D, et al. Prognostic value of metabolic parameters measured by first responders attending patients with severe trauma: associations with the new injury severity score and mortality. Emergencias. 2023;35:90-6.
23. Fokin AA, Wycech Knight J, Yoshinaga K, Abid AT, Grady R, Alayon AL, et al. Blunt cardiac injury in patients with sternal fractures. Cureus. 2022;14:e22841.
24. Cohn SM, Dubose JJ. Pulmonary contusion: an update on recent advances in clinical management. World J Surg. 2010;34:1959-70.
25. Elsaied Hussein MH, Fadl Mahmoud I, Ms Eita Y, Ahmed Aglan MA, Esmaiel MSA, Abdelshafy Ibrahim Farag G, et al. A prospective study of chest trauma scoring system as a morbidity and mortality predictor in patients with blunt chest trauma. Med J Islam Repub Iran. 2024;38:4.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.42744
Eurasian J Emerg Med. 2026;25: 147-55

SCAMPER-based First Aid Training: Effect on Students' Knowledge and Innovation-Oriented Attitudes

Fatih Cemal Tekin, Berke Yıldırım

Konya City Hospital, Clinic of Emergency Medicine, Konya, Türkiye

Abstract

Aim: This study aimed to evaluate the impact of the SCAMPER technique in first aid training on students' first aid awareness, development of safe-behavioral practices, and innovative product design thinking.

Materials and Methods: The study was conducted at Konya BİLSEM with 122 systematically sampled students, divided into two groups: one receiving traditional first aid training and the other instructed using the interactive SCAMPER technique. Pre- and post-training questionnaires assessed knowledge levels, anxiety about intervention, and attitudes toward equipment design. The traditional group was trained via classical presentations, while the SCAMPER group received training through interactive, creative exercises.

Results: Both groups showed a significant increase in first aid knowledge, adoption of safe-behavioral practices, and willingness to design first aid equipment after the training ($p<0.001$). The inclination to design innovative equipment was significantly higher in the SCAMPER group compared to the traditional group ($p<0.001$). Additionally, the majority of students expressed a desire for further first aid education (62.3%). SCAMPER-based training was associated with stronger innovation-oriented attitudes toward equipment design and interactive participation.

Conclusion: First aid education utilizing the SCAMPER technique is as effective as traditional methods in enhancing knowledge and fostering safe-behavioral practices, while offering additional benefits in stimulating innovative thinking regarding equipment design. Early first aid education incorporating SCAMPER is recommended to improve public health awareness and expand the reach of first aid practices within the community.

Keywords: Emergency medicine, first aid, education, educational models, innovativeness, early intervention, child behavior, experimental study

Introduction

First aid refers to the immediate care provided until medical staff arrive in unexpected events, accidents, acute health problems, or when encountering a life-threatening situation. The primary purpose of first aid is to eliminate danger, prevent the person's health from deteriorating further, and facilitate recovery. The most important distinction between first aid, emergency treatment, or emergency care and professional or emergency medical services is that "although first aid can often be delivered with minimal or improvised resources, the availability and appropriate use of basic first aid equipment may enhance safety and effectiveness in bleeding control." The timely and proficient administration of first aid using the correct techniques significantly contributes to the reduction of death and morbidity. Competent individuals are required for this. (1,2). Well-structured training programs are also necessary for individuals who are proficient in first aid practices.

There are different procedures for first aid training in various countries (3). Providing first aid training from an early age should be considered a strategic step for public health, not only for societal response within the emergency response system and the inclusion of the youngest individuals in this response, but also for the purpose of establishing safe-related behaviors behavioral. Therefore, including first aid lessons in the curriculum from elementary school onward, and repeating them in high school and university years, is beneficial in many ways (4). Because first aid is a need that can arise for every individual, initially in life-threatening situations, and sometimes in situations posing a smaller health risk. In addition, first aid application may be

Corresponding Author: Fatih Cemal Tekin MD, Konya City Hospital, Clinic of Emergency Medicine, Konya, Türkiye
E-mail: fatihcemaltekin@gmail.com **ORCID ID:** orcid.org/0000-0001-8410-5552
Cite this article as: Tekin FC, Yıldırım B. SCAMPER-based first aid training: effect on students' knowledge and ınnovation-oriented attitudes. Eurasian J Emerg Med. 2026;25: 147-55.

Received: 24.10.2025
Accepted: 06.01.2026
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

needed by anyone in the community, at any time, starting with the person themselves or their loved ones. Additionally, school is an environment where incidents requiring first aid are frequent, and these trainings can also raise awareness about accident prevention. Additionally, it will be easier to periodically repeat first aid training in formal education (5,6). In addition to this, although first aid doesn't require medical equipment, recognizing first aid materials is also an added benefit of these trainings. The creation of personal first aid kits in areas such as homes and vehicles, and the availability of first aid bags or cabinets in public spaces, are encouraged. Sometimes, however, this is a legal requirement (7-10). The advantage this brings is an increase in safe, hygienic, fast, accurate, and effective intervention. The basic principle regarding first aid supplies is that they should be easy to apply and readily available. However, these principles also highlight the need for ergonomic first aid supplies. This situation should not be overlooked in first aid training, and information about first aid equipment should also be provided (11,12).

In addition to traditional training methods, many new methods are being continuously tested and the results discussed regarding how first aid training will be delivered and the techniques to be used, utilizing the possibilities offered by technology. While the advancement of technology provides different opportunities for teaching and learning information, it also increases accessibility to knowledge (13). In addition to the goal of ensuring an adequate level of effectiveness in first aid training, it is expected to create the safe-behavioral practices, reduce avoidance of intervention during an event, and decrease nervousness or hesitant behavior. Another important topic that is perhaps overlooked in first aid training is the importance of training methods that also increase awareness of the introduction and use of first aid kits. This will increase the importance these individuals place on first aid kits, and will increase the attention given to having personal first aid kits or first aid supplies in public areas such as workplaces and schools. It will also facilitate the development of ergonomic, effective, and easy-to-use kits. In such an educational method, the components of brainstorming and interactive participation still maintain their importance, although certain limittaions have been reported (14-16). When examining educational methods, the SCAMPER technique, first proposed by Alex Osborn and developed by Robert F. Eberle, is one of the techniques used to facilitate creative thinking. This technique is named after the acronym of the steps: Substitute, Combine, Adapt, Modify/magnify/minify, Put to other uses, Eliminate, Reverse or rearrange. The advantage of the SCAMPER technique is that it provides an interactive learning environment by conducting brainstorming through specific steps, and prevents the topic from reaching a deadlock or from always revolving around the same ideas (17).

Within this context, the importance of first aid lessons in formal education for the development of first aid awareness, effective practices, and a safe-behavioral practices becomes evident. This topic is of particular interest from a medical perspective, including school health (18,19). In addition to this, another goal of the first aid training should be to develop ideas for the design of innovative products that can be used in the field of first aid and healthcare, which also relate to the field of medical engineering. This study aims to investigate the change in first aid awareness, s safe-behavioral practices, and innovative product design opinions regarding first aid areas and kits after training using the SCAMPER technique.

Materials and Methods

Study Design

The study was planned as an educational intervention study and examined how first aid awareness, safe-behavioral practices for bleeding, and innovative product design thinking in the field of first aid changed among groups when first aid training was delivered to two groups using the traditional method and the SCAMPER technique. The study was carried out at a Konya BİLSEM. In Türkiye, BİLSEM's are state-affiliated institutions where gifted students receive additional lessons in their areas of talent, alongside their formal education. This setting was selected due to the potential of students to engage in creative and innovation-oriented learning activities and this research could be more valuable in this regard for selected groups. According to the sample size calculation performed with the G* Power program for the research (test: Student's t test, effect size= 0.5 and $p<0.05$ and 85% power), it was found that a total of 118 people were needed for the two groups. The aim was to recruit 120 cases, with 60 for Group 1 and 60 for Group 2.

Permission was obtained from the KTO Karatay University Ethics Committee for Non-Pharmaceutical and Medical Device Research (desicion number 2024/015, date: 31.10.2024) for the research. Administrative leave was obtained from the institutions. Students' parents were informed with the help of the school administration and consent was obtained from the families. The trainings were held between 01.11.2024 and 30.11.2024.

Systematic Sampling and Student Selection

The systematic sampling method was used for sample selection. The complete list of BİLSEM students has been received. From this list, high school seniors and first and second-grade students were removed due to test anxiety and the intensity of their

programs, with the aim of preventing potential disagreements in reading and comprehension and reducing group heterogeneity. Then, planning was done to ensure an equal number of students were taken from each class. After registered students were ranked using student numbers for systematic sampling, at least 60 people were selected based on a random number table. Later, new students were selected in the same way to replace those who could not obtain permission from their parents or did not wish to participate. Group heterogeneity emerged during the election due to the internal dynamics of the center where the study was conducted. For students admitted to the center, there was a different number of students each year. Additionally, the lack of mandatory attendance at this center led to some students not participating and being unable to obtain permission from their families. To reduce group heterogeneity, students who scored high on the pre-test, those who received first aid training as part of the program, and students with a first-degree relative who is a healthcare worker were excluded from the study. Considering that the sample size calculated based on the two students' strong desire to participate in the education exceeded 5% of the total, both groups were increased to 61 people.

Education and Evaluation

Before the training, participants were asked questions such as, "Would you like to be a first responder?" and "Would you be worried if you encountered a person with bleeding who needed first aid?", and "Would you consider designing first aid equipment that could be used for bleeding patients?" In addition to this, a 15-question survey was administered, including a socio-demographic data. The questionnaire was developed by the authors through a literature review and reviewed for content validity by three experts in the field. However, no further psychometric testing (e.g., construct validity, reliability analysis) was conducted. The expected answer in the survey was coded as "1" point, while other answers were coded as "0" points. First aid awareness training was provided to the first group in groups of 20-20-21 people, using the classic presentation technique (first aid awareness training through PowerPoint and projector, equipment introduction, and practical applications), lasting 45 minutes. The training included an introduction to first aid equipment that can be used with the safe first aid behavior model in case of bleeding. The second group was taught using the SCAMPER technique in groups of 20-20-21 people each (first aid awareness training content, equipment introduction, and practical applications via PowerPoint and projector) (Figure 1). In both methods, the trainers covered the same topics in the same amount of time and by the same instructor. After the training, the 15-question survey was administered again, and the question "Would you like to receive more first aid training?" was also asked.

Statistical Analysis

The data obtained from the study results were analyzed using the Statistical Packages for the Social Sciences (SPSS) 18.0 Windows software package (SPSS Inc., Chicago, IL, US). When presenting the data, descriptive analyzes provided frequency data as numbers and percentages, while continuous numerical data were presented as mean ± standard deviation or median values. The normal distribution of the data was examined using the Shapiro-Wilk and Kolmogorov-Smirnov tests. The chi-square test

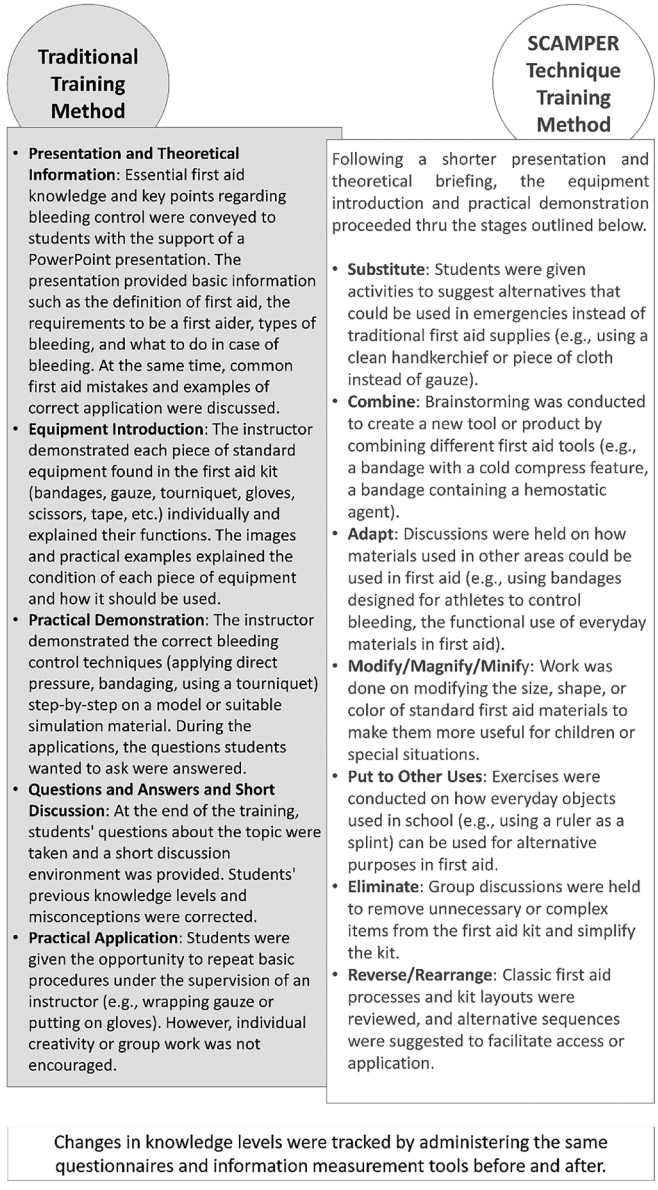


Figure 1. Implementation phases and illustrative examples of the traditional training method and the SCAMPER technique training method in first aid education

was used for comparing categorical data, the Wilcoxon test for comparing numerical data, and the Mann-Whitney U test. The Bonferroni correction was used to identify significant groups in the chi-square test with multiple eyes. Jamovi (version 2.6.26) software was used to determine effect sizes and CIs.

The level of statistical significance for all tests was considered to be $p<0.05$.

Results

The participants had a median age of 14 years, with a minimum age of 10 and a maximum age of 16. Of the participants, 62 (50.8%) were in primary and middle school, whereas 60 (49.2%) were in high school. The gender distribution consisted of 56% males (n=68) and 44% females (n=54). Of the participants, 73.8% (n=90) reported that they had not previously received first aid training. In response to the question, "Do you have sufficient knowledge about intervening in bleeding?" 82% of participants (n=100) indicated a lack of knowledge. Table 1 presents the changes in first aid knowledge levels following training, concerns regarding the application of first aid, and considerations for the design of first aid equipment and innovative products. The analysis revealed statistically significant differences in responses to the question "Do you have knowledge about first aid products/equipment that can be used in bleeding patients?" before and after the training ($p<0.001$, $X^2=133.954$). The observed difference originated from the group that responded affirmatively, the group that responded partially, and the group that responded negatively. The responses to the inquiry, "Would you consider designing first aid equipment for use on bleeding patients?" exhibited statistically significant differences pre- and post-training ($p<0.001$, $X^2=66.705$). A notable distinction emerged among the group that affirmed, the group that expressed uncertainty, and the group that denied.

The pre-training score of participants in the group trained with general first aid techniques was 6.13±1.88, while the post-training score was determined to be 10.07±1.67. It was found that the score for Safe Behavior and First Aid Awareness after the training was statistically significantly higher than before the training ($p<0.001$, effect size= 1.00). The data for the pre- and post-training comparison results are presented in Table 2.

It was found that the average of the scores participants received from the questionnaire administered before and after the training given using the SCAMPER technique was 6.16±2.07 before the training and 10.89±1.69 after the training. It was found that the score for Safe Behavior and First Aid Awareness after the training was statistically significantly higher than before the training ($p<0.001$, effect size= 1.00). The data for the pre- and post-training comparison results are presented in Table 3.

In the subgroup analyses for primary and secondary school students and high school students, the pre-training survey score was 5.85±2.14 for the primary and secondary school group, but the score for the high school group was 6.21±2.01. The mean

Table 1. Changes in Safe-Behavioral Practices and First Aid Awareness in Bleeding Before and After Training with Traditional First Aid Techniques (n=61)

		Pre-training		Pre-training		p
		n	%	n	%	
Would you like to be a first responder?	Yes	60	49.2	78	63.9	**p=0.02** $X^2=5.404$
	No	62	50.8	44	36.1	
Do you know how to get a first aid certificate?	Yes	8	6.6	118	96.7	**p<0.001** $X^2=198.504$
	No	114	93.4	4	3.3	
Do you know the conditions for being a first aider or who is called a first aider?	Yes	6	4.9	118	96.7	**p<0.001** $X^2=205.695$
	No	116	95.1	4	3.3	
Would you be concerned if you encountered someone who was bleeding and needed first aid?	Never	12	9.8	18	14.8	p=0.108 $X^2=6.076$
	Rarely	46	37.7	56	45.9	
	Often	50	41.0	42	34.4	
	Always	14	11.5	6	4.9	
Do you have any information about first aid products/equipment that can be used for bleeding patients?	Yes	16	13.0	106	86.9	**p<0.001** $X^2=133.954$
	Partially	72	59.0	14	11.5	
	No	34	27.9	2	1.6	
Would you consider designing first aid equipment that could be used on bleeding patients?	Yes	16	13.1	78	63.9	**p<0.001** $X^2=66.705$
	No idea	38	31.2	14	11.5	
	No	68	55.7	30	24.6	

post-training survey score rose to 9.72±2.98 for primary and secondary school pupils trained using the traditional technique, and to 10.88±1.68 for those trained with the SCAMPER technique. The average post-training survey score for high school students rose to 10.55±2.98 for those trained using the traditional method, and to 10.88±1.69 for those trained with the SCAMPER technique. No statistically significant difference was observed among the post-test scores of the primary, secondary, and high school groups following training using the SCAMPER approach (p=0.44, U=1709.0).

After the training, it was determined that 62.3% (n=76) of the participants answered "Yes" to the question "Would you like to take more first aid lessons or training?" A statistically significant difference was found between the idea of designing equipment and the increased demand for first aid training. Those considering designing new first aid equipment had a significantly higher demand for first aid lessons compared to other groups. Training provided using the SCAMPER technique (mean= 4.72±1.47) was statistically significantly associated with a greater inclination toward new equipment design compared to training provided using the traditional technique (mean= 3.93±1.58) [p<0.001, U=3178, mean difference= 1, 95% confidence interval (CI)= 1.00-1.50, effect size= 0.552]. Descriptive plots showing the changes in pre- and post-training survey scores with the training technique and the comparison of the two training techniques on these changes are presented in Figure 2.

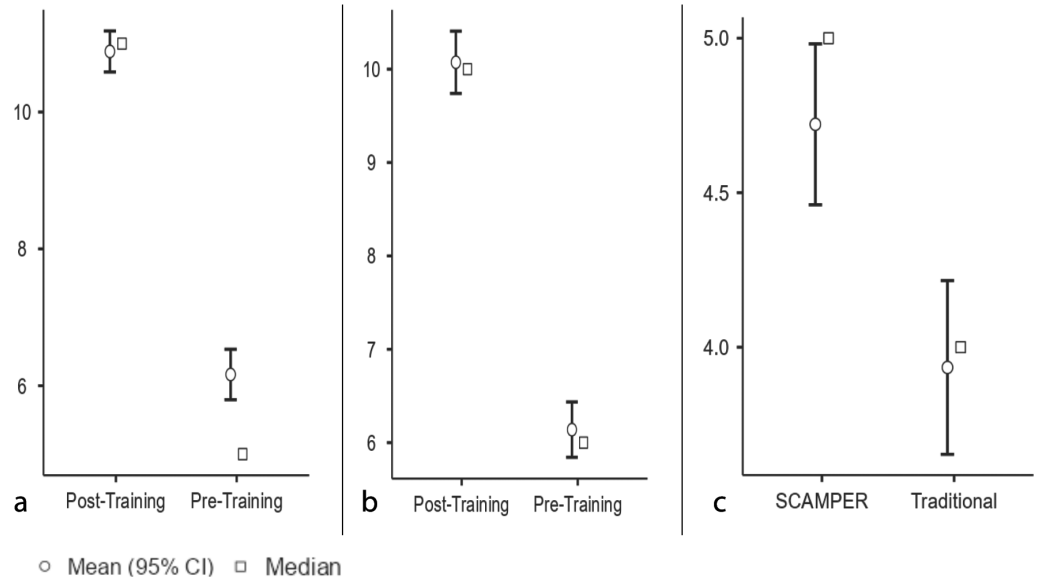


Figure 2. Descriptive plots. (a) Pre-test and post-test comparison of the SCAMPER technique. (b) Pre-test and post-test comparison of the traditional technique. (c) Comparison of the impact of SCAMPER and Traditional techniques on the idea of developing new first aid equipment, CI: Confidence interval

Table 2. Change in Safe Behaviors and First Aid Awareness in Bleeding Before and After Training with Traditional First Aid Techniques (n=61)

	Questionnaire score			p	Mean difference	95% CI	Effect size
	Mean ± SD	Median	Q^1-Q^3			Lower-upper	
Pre-training	6.13±1.88	5	5-7	**p<0.001** W=7381	4.00	3.50-4.50	1.00
Post-training	10.07±1.67	10	8-11				

CI: Confident interval, SD: Standard deviation, Q: Quartile, W: Wilcoxon test, SE difference= 0.143

Table 3. Pre- and Post-Training Changes in Safe Behaviors and First Aid Awareness in Bleeding Using the SCAMPER Technique (n=61)

	Questionnaire score			p	Mean difference	95% CI	Effect size
	Mean ± SD	Median	Q^1-Q^3			Lower-Upper	
Pre-training	6.16±2.07	5	5-8	**p<0.001** W=7503	4.50	4.50-5.00	1.00
Post-training	10.89±1.69	11	9-12				

CI: Confident interval, SD: Standard deviation, Q: Quartile, W: Wilcoxon test, SE difference= 0.133

Figure 3 present's several participants' concepts on the development and innovation of first aid products, generated throughout the training utilising the SCAMPER technique.

Discussion

According to the findings of this study, the training provided using traditional and SCAMPER techniques had a positive impact on increasing students' first aid awareness and ensuring safe behavior in bleeding situations. In the development of the idea of designing new equipment, the SCAMPER technique was found to be more effective.



Figure 3. Summary of ideas generated by participants during training using the SCAMPER technique, organized by stage

Studies on first aid knowledge levels emphasize the importance of starting first aid training at the primary school level, and even at the kindergarten level. It is reported that starting education at an early age also increases the expected benefits of first aid training repeated in later years (20,21). According to the results obtained in this study, 73.8% of primary-middle school and high school students stated that they had not received first aid training before, and 82.0% answered "no" to the question "Do you have sufficient knowledge about intervening in bleeding?" This situation not only highlights the lack of first aid training at an early age but also reveals the need for efforts in this area. Studies conducted in other countries regarding the pre-graduation first aid knowledge level also report being below the desired level, highlighting the need to work on this issue and provide training on first aid practices during education. It is known that in some countries, these trainings are offered with certification (22-24). Regarding first aid, there is an age limit in Türkiye. It is conceivable that similar practices may exist in other countries for legal reasons. Encouraging students to develop a positive attitude toward first aid and practice may be more important than the certificate program. However, a registration and tracking system can be established to ensure that all students receive similar training in this regard. It is also a fact that these awareness, understanding, and knowledge-raising practices for students will not provide the desired benefit to society without the participation of teachers and family support.

In this study, it was found that the level of knowledge about first aid and the conditions for being a first aider significantly increased after the training provided. One of the most striking points is the positive change in the desire to be a first responder. This output not only demonstrates that first aid training should be part of community-based awareness-raising efforts, but also supports studies (2,4,6,20,21,25) that emphasize the importance of providing first aid training at an early age. There is also a continued need for developments in the most accurate first aid training methods for this age group. Because general first aid training and community-based practices are mostly single-session, non-repetitive training. However, research is increasing that suggests user experiences and interaction-focused training methods will be more beneficial for learners in first aid training. Virtual reality, game-based or mobile applications, and many different methods focused on personalization and gamification are still being tested and positive results are being reported (25-27). In this study, an intensive method was used in terms of interaction and interactive communication, and findings were reached that support these results. However, the fact that there was no significant change in the responses to the question of whether they would hesitate to intervene in a bleeding patient after the training indicates that practitioners still have some

hesitation about taking action. In many publications, it is seen that this is a general problem affecting other age groups as well (28-30). Although it is thought that this problem can be somewhat reduced by increasing practical training, and the results indicate the need for new studies on the causes of this situation.

Before training, 13% of students stated they had sufficient knowledge of the equipment that could be used on a bleeding patient and most people thought they didn't have enough knowledge on this topic. This result is also related to the topic discussed in the first paragraph, as it could be an indicator of the lack of early first aid training. Additionally, this could be due to insufficient introduction of the kits during first aid training, or an overemphasis on the "without the need for any materials" aspect of first aid's definition. This is actually a situation that highlights the distinction between first aid and emergency care. However, using a kit or equipment for first aid is beneficial in many ways and is encouraged in many countries. This point should be carefully emphasized during training (10,31,32). It was observed that the group who said they had knowledge about the equipment that could be used on a bleeding patient after the training was statistically significantly larger.

The evaluation of response scores from the 15-question survey presented to participants revealed that both the standard method and the SCAMPER technique resulted in a statistically significant enhancement in students' awareness of first aid and their attitudes towards the safe behaviour model for wounds. This is significant as it demonstrates that the SCAMPER technique is equally effective as traditional methods in first aid awareness training. This serves as evidence that the SCAMPER technique is applicable in first aid training. The absence of a statistically significant difference in the post-training test survey results between the elementary and secondary school age groups and the high school age group indicates that the SCAMPER technique exhibits comparable effectiveness throughout these subgroups.

According to the research results, there was also a significant increase in the participants' thinking about designing new equipment compared to before the training. Regarding increasing students' awareness of designing first aid equipment, the SCAMPER technique was more effective than the traditional method. While research on new equipment in the field of first aid is ongoing, children's first aid practices will likely require more equipment usage than adults. For example, the basic first aid intervention for bleeding is direct pressure (33). It could be difficult for him to apply and maintain this pressure during the period when there is no other savior beside the children, or until the request for help is met. This equipment differentiation will undoubtedly be even more pronounced in critical applications such as basic life support. The accessibility of first aid kits, the clarity and applicability of instructions during use, are also important issues (34,35). While recommendations for the composition of materials used for children in first aid kits are still ongoing (10), designing equipment for a child first responder might be premature at this point and this could be the subject of another research project. However, first aid training will undoubtedly be very important for raising awareness of accidents and for developing safe behaviors in emergency situations. Also, the idea of designing equipment is undoubtedly an idea that will be used in the future. Therefore, as in this study, using the SCAMPER technique in first aid training for students who could be in the target group in terms of equipment design, and who are considering fields such as engineering or medical engineering, could be beneficial. In addition, using the SCAMPER technique in first aid training can have a positive impact on first aid kit innovation, as it can generate ideas for using everyday materials as first aid tools in other groups as well.

When examining publications that emphasize the importance of disseminating first aid knowledge within society and reiterate this need, it is evident that research on how to approach this issue is still ongoing. It would not be the correct approach to consider the educational method's subject matter, instructor, environment, and materials independently of the target group. This situation leads us to believe that the expectation of a single and most correct educational method does not reflect reality. The correct training method will also differ depending on the variables. Current discussions and research focus on the impact of technology use and the increasing online methods it will create. In addition, the lack of structured tests that can fully measure the results of first aid training poses a significant problem for researchers in determining the correct training methods for first aid (16,26). The focus of this study is that different first aid methods can yield different results in designing innovative first aid products. In this regard, the SCAMPER technique stands out as an attention-grabbing educational method. Publications related to the development of first aid equipment and kits are not common in the literature. This could be related to the economic and commercial aspects of this issue. The importance of health interventions performed until emergency medical assistance arrives has been highlighted many times. Although these applications focus more on basic life support, there is a need to attract researchers' interest in the topics of first aid and equipment. It is extremely important for school children to acquire the concept of first aid with this awareness for their future perspective. The fact that participants in the study expressed a greater demand for first aid training also indicates that there is a willing party for this, suggesting that this opportunity should not be missed. However, instilling this idea should not be limited to school age. In addition to the idea of facilitating access to

equipment for rescuers from the public, necessary steps should be taken to make basic first aid tools ergonomic and easy to use.

Study Limitations

This study was conducted with a relatively small number of students. Although studies with similar or even smaller sample sizes can be found in the literature, the fact that the study was conducted at only one center requires careful interpretation of the results in terms of generalizability. The study focused on the use of the SCAMPER technique in first aid training. Additionally, since the center where the study was conducted is a school that accepts talented students selected from many different schools, and the class and age distribution were not balanced, more detailed analyzes could not be performed for subgroups (especially by age level). Other reasons for the wide age range of participating students were that groups did not have to come to the center (attendance was optional) and that students came to the center on different days and at different times. This limitation was partially addressed through subgroup analysis, which showed that students in primary, middle, and high school benefited similarly from the SCAMPER technique. In addition, heterogeneity and limited sample size have led to limitations in the interpretation of the findings. However, the pre-test found generally similar levels of first aid knowledge for both groups. In some students, a higher level of awareness was observed, influenced by their families, personal interests, and education. To minimize comment errors, factors such as outliers during sampling and the presence of a healthcare worker in the family were considered exclusion criteria. We believe this can be considered a factor in reducing heterogeneity and contributing to the accurate interpretation of findings.

Another important limitation is that the 15-item questionnaire used in this study was developed by the authors. Although reviewed by field experts, no formal validity or reliability (psychometric) analysis was conducted on the survey. Therefore, the psychometric properties of the survey instrument have not been definitively established, which could affect the generalizability and robustness of the findings. In future studies, it is recommended to use measurement tools with proven validity and reliability and to conduct psychometric evaluations.

Conclusion

This study emphasizes the need of early first aid training with the SCAMPER technique for innovative equipment design and creative thinking. This age group has a poor level of first aid knowledge and awareness, and there is a need to improve this condition in order to strengthen the effectiveness of the emergency response system in their future lives as well as promote public health awareness and levels. However, the readiness of participating students to receive more first aid classes suggests that this group is receptive to a prospective early first aid training campaign, waiting for a chance, and eager to fill gaps in first aid knowledge. In this study, it was discovered that the SCAMPER technique is as effective as the traditional way in boosting first aid awareness and knowledge levels, as well as in developing a safe behavior model in emergency scenarios. This shows that incorporating this strategy into first aid training is appropriate. Furthermore, the SCAMPER technique, which is well-known for its benefits in brainstorming and interactive learning, can be applied in first aid training and equipment introduction to increase awareness of first aid kits. Furthermore, encouraging students to design first aid equipment and create creative goods in this field can emphasize the benefits of this training method, particularly in certain groups.

Ethics

Ethics Committee Approval: Permission was obtained from the KTO Karatay University Ethics Committee for Non-Pharmaceutical and Medical Device Research (desicion number 2024/015, date: 31.10.2024) for the research.

Informed Consent: Students' parents were informed with the help of the school administration and consent was obtained from the families.

Footnotes

Authorship Contributions

Concept: F.C.T., B.Y., Design: F.C.T., B.Y., Data Collection or Processing: F.C.T., B.Y., Analysis or Interpretation: F.C.T., Literature Search: F.C.T., Writing: F.C.T., B.Y.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Dağ R. Temel İlk yardım uygulamaları eğitim kitabı. İstanbul: Beta; 2018.
2. Parada-Espinosa JM, Ortega-Gómez S, Ruiz-Muñoz M, González-Silva J. Teaching methodologies for first aid in physical education in secondary schools: a systematic review. Healthcare (Basel). 2025;13:1112.
3. Ygiyeva D, Pivina L, Messova A, Omarov N, Batenova G, Jamedinova U, et al. Evaluating the effectiveness of a first aid training programme for individuals without a background in medical education. Adv Med Educ Pract. 2024;15:773-82.
4. Kuvaki B, Özbilgin Ş. School children save lives. Turk J Anaesthesiol Reanim. 2018;46:170-5. Epub 2017 Jun 1.
5. Correia LFR, Feitosa AF, Apolinario MVF, Sampaio EF, Braga ST, Pinheiro WR. The importance of teaching and learning first aid skills to lay people: an integrative review. R Pesq Cuid Fundam. 2024;16:e11605.

6. de Melo Alves Silva LC, Alves IL, dos Santos KVG, da Silva TTM, da Silva Leal KC, et al. First aid teaching for schoolchildren: Scoping review. International Journal of Educational Research Open. 2023;5:100305.
7. American Red Cross. Make a First Aid Kit. Available from: https://www.redcross.org/get-help/how-to-prepare-for-emergencies/anatomy-of-a-first-aid-kit.html?srsltid=AfmBOooyvZ-4nfp_0yudeUuu38xwGIvrHm7ecKc-bIYusIb2aBONxfe1.
8. British Red Cross. First aid kit contents: Essential items for your workplace. 2025 (cited 2025 October 22). Available from: https://www.redcrossfirstaidtraining.co.uk/first-aid-supplies-blogs/first-aid-kit-contents-essential-items-for-your-workplace/
9. Resmî Gazete. İş sağlığı ve güvenliği kanunu. Available from:https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=6331&MevzuatTur=1&MevzuatTertip=5
10. Oliveira ATB. Worldwide regulation of the medical emergency kit and first aid kit. Aerosp Med Hum Perform. 2024;95:321-6.
11. Wen D, Wan L, He H, Zhong Q, Jiang Q, Yang X, A cross-sectional survey of first-aid kit equipment in a family in Sichuan, China. BMC Public Health. 2024;24:1829.
12. Mahpar N, Jumadi NA. Development of first aid kit with ıntuitive touch screen display and voice ınstruction. Journal of Electronic Voltage and Application. 2021;2:1-10.
13. Çaresiz B, Berkalp H, Kılınç R, Akçi K, Karakaya C. Eğitim yönetimi modelleri ve uygulamaları: Güncel yaklaşımlar. International QMX Journal. 2024;3:394-402.
14. Paulus PB, Baruah J, Kenworthy J. Brainstorming: how to get the best ideas out of the "group brain" for organizational creativity. Handbook of Organizational Creativity. Cambridge: Academic Press. 2023. p. 373-89.
15. Wieland B, de Wit J, de Rooij A. Electronic brainstorming with a chatbot partner: a good idea due to increased productivity and idea diversity. Front. Artif. Inte. 2022;23:e880673.
16. Minna S, Leena H, Tommi K. How to evaluate first aid skills after training: a systematic review. Scand J Trauma Resusc Emerg Med. 2022;30:56.
17. Serrat O. The SCAMPER Technique. In: Serrat O, editors. Knowledge solutions: tools, methods, and approaches to drive organizational performance. Singapore: Springer: 2017. p. 311-314.
18. Yeşildağ AY, Akbulut, Uğurluoğlu Aldoğan E. Örgün eğitim programında sağlık eğitiminin önemi: içerik analizine dayalı bir araştırma. Eğitim ve Bilim. 2024;49: 233-54.
19. Kurt G, Serdaroğlu HU, Özcan C. Türkiye'de okul sağlığı araştırmaları. Arşiv Kaynak Tarama Dergisi. 2024;33:198-209.
20. Tse E, Plakitsi K, Voulgaris S, Alexiou GA. First aid training for children in kindergarten: a pilot randomized control study. Children. 2022;9:1626.
21. Tse E, Plakitsi K, Voulgaris S, Alexiou GA. The role of a first aid training program for young children: a systematic review. Children. 2023;10:431.
22. Ng SN, Tang LK, Leung CK, Cheng CY, Cheung MS, Lam YY, et al. Knowledge and training preference of standard first aid among undergraduates in Hong Kong: a cross-sectional survey. Human Behavior and Emerging Technologies. 2023;30:172-8.
23. Dahal G, Vaidya P. J Nepal Health Res Counc. Knowledge of first aid in school students and teachers. J Nepal Health Res Counc. 2022;20:96-101.
24. García-Blaya JÁ, Abraldes JA, Vaquero-Cristóbal R. Assessment of first aid knowledge at different stages of education. Healthcare (Basel). 2025;13:1507.
25. Farooq F, Ishaq K, Alvi A, Rosdi F, Nawaz NA. Transforming healthcare training: personalized and gamified first aid education. Human Behavior and Emerging Technologies. 2025;2025:26.
26. Luo J, Zheng K, Hong W. Public first aid education model design study based on user experience. Front Public Health. 2023;18:1286250.
27. Santana RCB, Oliveira MMC, Rios AAS, Galindo Neto NM. Mobile application for first aid to children: a technological innovation in schools. Rev Bras Enferm. 2025;78:e20240057.
28. Kusumawati HI, Sutono Alim S, Achmad BF, Putri AF. Factors associated with willingness to perform basic life support in the community setting in Yogyakarta, Indonesia. Australas Emerg Care. 2023;26:303-7.
29. Zhong Z, Li W, Zhou L, Liu L. Factors influencing the performance of cardiopulmonary resuscitation by lay rescuers: a qualitative study based on the theory of planned behavior. PLoS One. 2025;20:e0327439.
30. Katona Z, Tarkó K, Berki T. First aid willingness questionnaire for schoolchildren: an exploratory factor analysis and correlation study. Children. 2022;9:955.
31. Ge P, Zhang J, Lyu K, Niu Y, Li Q, Xiong, P, et al. The current status and factors related to the preparation of home first-aid kits in China. Front Public Health. 2022;10:1036299.
32. Kulkarni AJ, Thiagarajan AB, Ogana SO, Okwiri DA, Arudo J, Smith N, et al. Evaluating lay first responder (LFR) first aid kit supplies usage and appropriateness in Western Kenya. Pan Afr Med J. 2024;48:169.
33. Schroll R, Smith A, Martin MS, Zeoli T, Hoof M, Duchesne J, et al. Stop the bleed training: rescuer skills, knowledge, and attitudes of hemorrhage control techniques. J Surg Res. 2020;245:636-42.
34. Steins K, Goolsby C, Grönbäck AM, Charlton N, Anderson K, Dacuyan-Faucher N, et al. Recommendations for placement of bleeding control kits in public spaces-a simulation study. Disaster Med Public Health Prep. 2023;17:e527.
35. Alkinani MH, Almazroi AA, Jhanjhi N, Khan NA. 5G and IoT based reporting and accident detection (RAD) system to deliver first aid box using unmanned aerial vehicle. Sensors. 2021;21:6905.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.95826
Eurasian J Emerg Med. 2026;25: 156-65

Artificial Intelligence Applications in Oral and Maxillofacial Surgery: A Bibliometric and Science Mapping Analysis of Global Research Trends (2000-2025)

Ceyda Gizem Topal[1], Mustafa Kenan Hürmüzlü[2], Murat Ünal[3], Nursel Arpay[4], Osman Yalçın[1], Ömer Onur Gümüş[5]

[1]Yozgat Bozok University Faculty of Dentistry, Department of Oral and Maxillofacial Radiology, Yozgat, Türkiye
[2]Yozgat Bozok University Faculty of Dentistry, Department of Oral and Maxillofacial Surgery, Yozgat, Türkiye
[3]Yozgat Bozok University Faculty of Dentistry, Department of Endodontics, Yozgat, Türkiye
[4]Antalya Bilim University Faculty of Dentistry, Department of Oral and Maxillofacial Radiology, Antalya, Türkiye
[5]Independent Researcher

Abstract

Aim: To map the scientific landscape of artificial intelligence (AI) applications in oral and maxillofacial surgery (OMFS) over the last 25 years using bibliometric and science-mapping methods.

Materials and Methods: Publications related to AI and OMFS were retrieved from the Web of Science Core Collection. A comprehensive search was performed, and the complete set of 2153 records was exported. After restricting the timespan to 2000-2025 and limiting the document type to articles and reviews, 1955 documents were included. Bibliometric indicators were calculated, and science maps were generated in VOSviewer for co-authorship, co-occurrence, citation, and co-citation analyses.

Results: The annual number of publications increased markedly after 2015, with a steep rise from 2020 onwards and a peak of 275 publications in 2025. Overall, 1678 articles (87.8%) and 232 reviews (12.2%) were identified, totaling 38648 citations (mean 20.2 citations per document). The most productive journals were the Journal of Craniofacial Surgery, the Journal of Cranio-Maxillofacial Surgery, and the International Journal of Oral and Maxillofacial Surgery. Keyword co-occurrence analysis revealed major thematic clusters related to orthognathic surgery, mandibular reconstruction, computer-assisted surgery, virtual surgical planning, 3D printing, and deep learning.

Conclusion: AI research at the interface of OMFS has expanded rapidly between 2000 and 2025, particularly in AI-assisted surgical planning, anatomical segmentation for operative workflows, and computer-assisted surgery. These findings highlight how AI-supported technologies are increasingly being investigated for integration into contemporary clinical workflows in OMFS and provide guidance for clinicians regarding future implementation and validation priorities. In addition to clinical insights, these findings provide an overview of the AI research landscape in OMFS, offering guidance for future algorithm development, validation strategies, and interdisciplinary research directions.

Keywords: Artificial intelligence, bibliometric analysis, oral and maxillofacial surgery, science mapping, surgical planning

Introduction

Over the past 10 years, advances in machine learning and deep learning, along with the increased availability of digital surgical data, have driven growing interest in artificial intelligence (AI) in oral and maxillofacial surgery (OMFS). The growing importance of AI in OMFS has been highlighted by recent narrative and systematic studies, particularly in surgical planning, anatomical analysis, and clinical decision-making. (1-4).

Among the various clinical applications, orthognathic surgery has emerged as a key area for integrating digital and AI-assisted technologies. Virtual surgical planning, simulation, and patient-specific surgical guides have been widely investigated as tools to

Corresponding Author: Ceyda Gizem Topal MD, Yozgat Bozok University Faculty of Dentistry, Department of Oral and Maxillofacial Radiology, Yozgat, Türkiye
E-mail: cyd.hsvr@gmail.com **ORCID ID:** orcid.org/0000-0001-7401-418X

Cite this article as: Topal CG, Hürmüzlü MK, Ünal M, Arpay N, Yalçın O, Gümüş ÖO. Artificial intelligence applications in oral and maxillofacial surgery: A Bibliometric and Science Mapping Analysis of Global Research Trends (2000-2025). Eurasian J Emerg Med. 2026;25: 156-65.

Received: 26.12.2025
Accepted: 06.01.2026
Published: 26.01.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

improve surgical accuracy and predictability (5-9). In parallel, deep learning–based methods have been introduced to automate tasks such as mandibular segmentation and identification of critical anatomical structures, which are essential components of contemporary OMFS workflows (10-14).

These methods enhance surgical planning and operative decision-making rather than relying solely on stand-alone radiological evaluations, even though they often rely on imaging-derived data. This differentiation highlights the essentially surgical focus of AI applications in OMFS and distinguishes them from studies in diagnostic imaging.

Several authors have discussed the broader implications of AI in dentistry and healthcare, in addition to procedure-specific studies. These contributions offer crucial conceptual and methodological frameworks for comprehending the creation, verification, and constraints of AI-based systems in clinical practice (16-18). More recently, OMFS-specific systematic reviews have synthesised evidence on AI applications across surgical subspecialties, highlighting the field's existing methodological shortcomings and potential advantages (19).

Despite the growing volume of literature, existing reviews primarily offer qualitative or application-focused perspectives and provide limited insight into publication trends, collaborative research structures, and the thematic evolution of AI-related OMFS research over time. Bibliometric and science mapping approaches are well-suited to address these gaps by quantitatively assessing research output, identifying influential contributors, and visualising emerging research themes.

Beyond its clinical relevance, systematically mapping the structure, evolution, and thematic focus of AI research in OMFS is essential to guide future algorithm development, foster interdisciplinary collaboration, and identify methodological gaps within the AI research landscape (16-18).

Therefore, the present study aimed to conduct a comprehensive bibliometric and science-mapping analysis of AI applications in OMFS published between 2000 and 2025.

Materials and Methods

Study Design

This study followed a bibliometric and science-mapping design and was conducted in accordance with reporting recommendations for bibliometric analyses. Because only published literature was analysed and no individual patient data were used, ethical approval was not required.

Data Source and Search Strategy

The Web of Science (WoS) Core Collection was chosen as the data source for its strict indexing guidelines and suitability for citation-based analysis. A thorough search was conducted to find articles on AI applications in OMFS. OMFS and dental imaging terms were merged with free-text terms pertaining to deep learning and AI in the search approach. Excluded were studies with a primary focus on dentomaxillofacial radiography that had no direct bearing on surgical planning, intervention, or results in OMFS. A representative Boolean query was:

Topic search= [("AI " or "deep learning" or "machine learning" or "neural network*" or "radiomics")] and ("oral and maxillofacial" or "maxillofacial surgery" or "orthognathic" or "mandibular reconstruction" or "dentofacial deformit*" or "dental implant*" or "cone-beam computed tomography" or " CBCT" or "panoramic radiograph*").

The initial search of the WoS Core Collection retrieved 2153 records with no restrictions. All records were exported in plain-text format with complete records and cited references.

Study Selection and Data Cleaning

In a first screening step, the timespan was restricted to publications from 1 January 2000 to 31 December 2025. Publication counts for 2025 reflect partial-year data, as records were retrieved on 9, 2025, before the calendar year ended. Only items indexed as "article" or "review" were retained; conference proceedings, editorial materials, letters, and corrections were excluded by using WoS document-type filters. After this restriction and the exclusion of records, 1955 documents remained and were included in the quantitative analyses.

To prepare the data for science-mapping analyses, author names, institutional affiliations, and keywords were cleaned and normalised. Variants of the same term (e.g., "computer-aided surgery" and "computer-assisted surgery") were merged, spelling errors were corrected, and singular and plural forms were unified where appropriate. Author keywords (DE field) and Keywords Plus (ID field) were combined to generate a unified keyword list. Data cleaning was performed using spreadsheet tools and custom scripts before importing the dataset into VOSviewer. Conceptually overlapping terms such as computer-aided surgery, computer-assisted surgery, and computer-aided design (CAD)/computer-aided manufacturing (CAM) were merged under unified keyword categories prior to frequency analysis. Imaging-related terms were included to ensure comprehensive retrieval of AI studies supporting surgical planning and workflows, while studies focusing solely on diagnostic radiology without surgical relevance were excluded during data cleaning. Following data

cleaning procedures and the application of predefined inclusion criteria, 1.955 publications were retained for bibliometric and science mapping analyses.

Bibliometric Indicators

Descriptive bibliometric indicators were calculated for the overall dataset, including the annual number of publications, document types, total and average citations per document, and productivity rankings for countries, institutions, authors, and journals. Citations were counted as of the WoS export date.

Statistical Analysis

Science-Mapping Analysis

Science maps were generated using VOSviewer (version 1.6.20; Centre for Science and Technology Studies, Leiden University, The Netherlands). Co-authorship analyses were performed at the author, institution, and country levels using complete counting. Co-occurrence analyses were conducted for author keywords; only keywords with a minimum of five occurrences in the dataset were included in the maps. Citation and co-citation analyses were carried out for journals and individual documents. Normalisation of similarities was based on the association strength method, and clusters were identified using VOSviewer's default clustering algorithm. Overlay visualisation was used to explore temporal trends, with node colour representing the average publication year of items.

Results

Descriptive Overview of the Dataset

A total of 1955 WoS-indexed documents on AI applications in OMFS were published between 2000 and 2025. Of these, 1720 (87.9%) were original research articles and 235 (12.1%) were review papers. The cumulative number of publications increased modestly in the early 2000s and then accelerated from approximately 2015 onwards, with a marked rise after 2020, peaking at 275 in 2025. (Figure 1) The total number of citations received by the included documents was 38648, corresponding to an average of 20.2 citations per document. Most publications were written in English (97.4%), with only a limited number published in other languages.

The country collaboration network is visualised in Figure 2, revealing dense collaborative clusters among high-output countries and indicating that international co-authorship plays a substantial role in advancing research on AI in OMFS.

Most Highly Cited Publications

Table 1 lists the 20 most highly cited documents in the dataset. These publications predominantly address 3D printing, computer-assisted and image-guided surgery, virtual surgical planning, and mandibular reconstruction, which together form the historical core of digital OMFS.

Most Productive Journals

The most productive journals in this field were Journal of Craniofacial Surgery (n=145), Journal of Cranio-Maxillofacial Surgery (n=138), International Journal of OMFS (n=123), and Journal of Oral and Maxillofacial Surgery (n=108). Several multidisciplinary and engineering journals, such as Applied Sciences and Scientific Reports, also contributed substantially to the literature, reflecting the interdisciplinary nature of AI research in OMFS. The 10 most productive journals publishing AI-related OMFS research are summarised in Table 2. The dominance of core OMFS journals highlights that AI-related research is being increasingly disseminated within speciality-specific surgical outlets rather than general radiology or engineering journals.

Keyword Co-occurrence

Author keywords and Keywords Plus were combined for co-occurrence analysis. Table 3 presents the 20 most frequent keywords in the dataset. Orthognathic surgery, computer-assisted surgery, AI and mandibular reconstruction were among the most common terms, together with virtual surgical planning and CAD/CAM. Imaging-related keywords such as CBCT appeared predominantly in the context of surgical planning, navigation, and outcome assessment in OMFS, rather than as stand-alone radiological applications. "Accuracy" was among the most frequently occurring keywords, particularly with respect to surgical outcome prediction and anatomical segmentation tasks in OMFS. The keyword co-occurrence network is presented in Figure 2, demonstrating several well-defined thematic clusters centred on AI applications in OMFS, including surgical planning, deep learning-based segmentation, and computer-assisted surgical workflows.

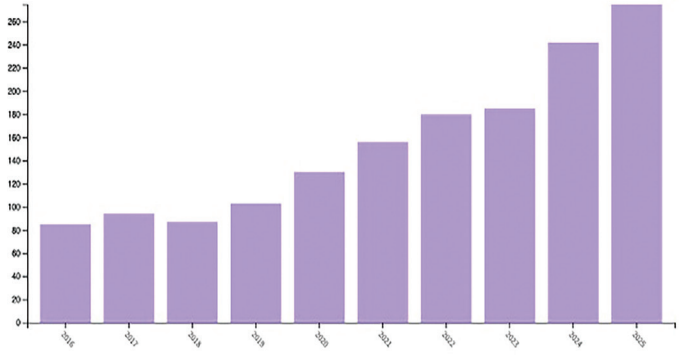


Figure 1. Annual publication trend (2000-2025) Data for the year 2025 represent a partial dataset collected up to December 9, 2025, and therefore do not reflect a complete calendar year

The overlay visualisation shown in Figure 3 illustrates the temporal evolution of research topics, indicating a shift from early computer-assisted surgical approaches toward a more recent emphasis on deep learning, virtual surgical planning, and data-driven decision-support systems in OMFS.

Geographical Distribution of Publications

The geographical distribution of publications revealed that research on AI in OMFS was dominated by a limited number of countries (Table 4). The People's Republic of China and the United States were the most productive countries, followed by Germany, Italy, and South Korea. Other major contributors included Japan, France, and the Netherlands, indicating strong research activity across East Asia and Western Europe. The country collaboration network is visualised in Figure 4, revealing dense collaborative clusters among high-output countries and indicating that international co-authorship plays a substantial role in advancing research on AI in OMFS.

Institutional Productivity

Institutional analysis demonstrated that a small number of academic and clinical centres accounted for a substantial proportion of the published literature (Table 5). Shanghai Jiao Tong University and Peking University emerged as the most productive institutions, followed by several major universities and university hospitals from East Asia, Europe, and North America. This distribution reflects the concentration of AI-related OMFS research within high-volume tertiary care and research-orientated institutions.

Author Productivity

Analysis of author productivity showed that a limited group of researchers contributed disproportionately to the literature (Table 6). Gellrich NC, Lo LJ, and Marchetti C were among the most productive authors, each contributing more than 20 publications. The remaining highly productive authors demonstrated comparable publication outputs, suggesting the

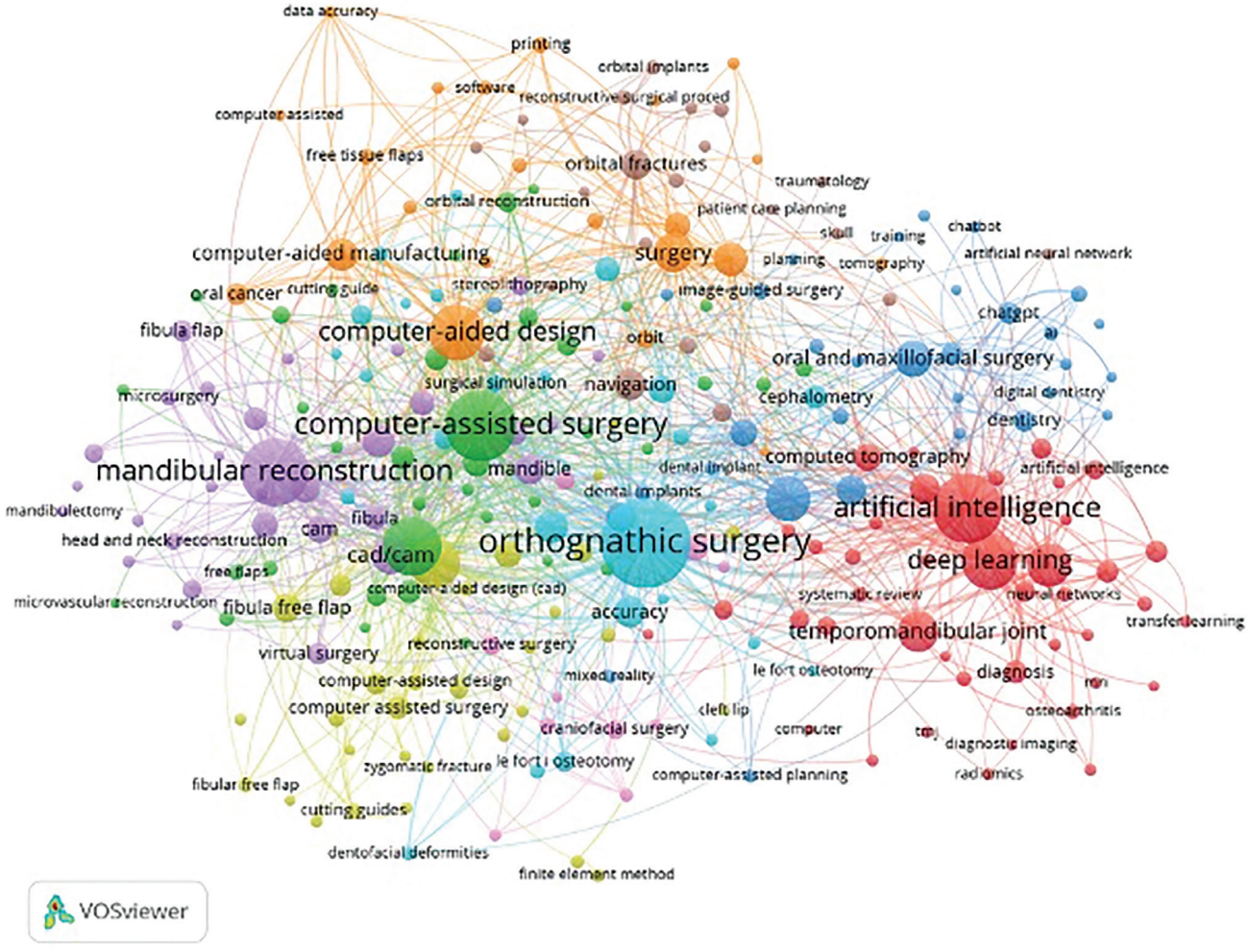


Figure 2. Keyword co-occurrence network generated using VOSviewer. Colours represent thematic clusters, and node size indicates the relative frequency of each keyword in the dataset

Table 1. The 20 most highly cited publications on artificial intelligence applications in oral and maxillofacial surgery (2025-2000)

Rank	First author	Year	Title	Journal	Document type	Citations
1	Tack, Philip	2016	3D-printing techniques in a medical setting: a systematic literature review	Biomedical Engineering Online	Review	738
2	Warnke, PH	2004	Growth and transplantation of a custom vascularised bone graft in a man	Lancet	Article	537
3	Martelli, Nicolas	2016	Advantages and disadvantages of 3-dimensional printing in surgery: a systematic review	Surgery	Review	460
4	Hsu, Sam Sheng-Pin	2013	Accuracy of a computer-aided surgical simulation protocol for orthognathic surgery: a prospective multicenter study	Journal of Oral and Maxillofacial Surgery	Article	321
5	Cohen, Adir	2009	Mandibular reconstruction using stereolithographic 3-dimensional printing modelling technology	Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontology	Article	281
6	Tian, Yueyi	2021	A review of 3D printing in dentistry: technologies, affecting factors, and applications	Scanning	Review	275
7	Hirsch, David L.	2009	Use of computer-aided design and computer-aided manufacturing to produce orthognathically ideal surgical outcomes: a paradigm shift in head and neck reconstruction	Journal of Oral and Maxillofacial Surgery	Article	267
8	Chae, Michael P.	2015	Emerging applications of bedside 3D printing in plastic surgery	Frontiers in Surgery	Review	257
9	Malik, Hammad H.	2015	Three-dimensional printing in surgery: a review of current surgical applications	Journal of Surgical Research	Review	250
10	Plooij, Joanneke M.	2011	Digital three-dimensional image fusion processes for planning and evaluating orthodontics and orthognathic surgery. A systematic review	International Journal of Oral and Maxillofacial Surgery	Review	221
11	Hoang, Don	2016	Surgical applications of three-dimensional printing: a review of the current literature & how to get started	Annals of Translational Medicine	Review	217
12	Hanasono, Matthew M.	2013	Computer-assisted design and rapid prototype modelling in microvascular mandible reconstruction	Laryngoscope	Article	215
13	Widmann, G	2006	Accuracy in computer-aided implant surgery - a review	International Journal of Oral & Maxillofacial Implants	Review	206
14	Block, Michael S.	2017	Implant placement accuracy using dynamic navigation	International Journal of Oral & Maxillofacial Implants	Article	206
15	Hassfeld, S	2001	Computer-assisted oral and maxillofacial surgery - a review and an assessment of technology	International Journal of Oral and Maxillofacial Surgery	Review	203
16	Ewers, R	2005	Basic research and 12 years of clinical experience in computer-assisted navigation technology: a review	International Journal of Oral and Maxillofacial Surgery	Review	201
17	Eggers, G.	2006	Image-to-patient registration techniques in head surgery	International Journal of Oral and Maxillofacial Surgery	Review	189
18	Stokbro, K.	2014	Virtual planning in orthognathic surgery	International Journal of Oral and Maxillofacial Surgery	Article	187
19	Plooij, J. M.	2009	Evaluation of reproducibility and reliability of 3D soft tissue analysis using 3D stereophotogrammetry	International Journal of Oral and Maxillofacial Surgery	Article	183
20	Antony, Anuja K.	2011	Use of virtual surgery and stereolithography-guided osteotomy for mandibular reconstruction with the free fibula	Plastic and Reconstructive Surgery	Article; Proceedings Paper	182

presence of several active research groups rather than a single dominant author.

Discussion

This bibliometric and science-mapping analysis demonstrates a steep and accelerating increase in AI-related publications in OMFS over the last 25 years. The surge in output after 2015 parallels the broader diffusion of deep learning frameworks and accessible GPU computing in medicine (1,3,15). Recent narrative reviews similarly emphasise that AI has transitioned from experimental proof-of-concept work to an integral component of digital OMFS workflows (1-4).

Geographical Distribution and International Collaboration

Analysis of publication output demonstrated that research activity related to AI in OMFS is concentrated in a limited number of countries. China emerged as the most prolific contributor, followed by several European countries, including Germany and Italy. Similar geographical patterns have been described

Table 2. The most productive journals publishing AI-related OMFS research (2025-2000)

Rank	Journal	Number of Documents
1	Journal of Craniofacial Surgery	145
2	Journal of Cranio-Maxillofacial Surgery	138
3	International Journal of Oral and Maxillofacial Surgery	123
4	Journal of Oral and Maxillofacial Surgery	108
5	Journal of Stomatology Oral and Maxillofacial Surgery	44
6	Journal of Clinical Medicine	43
7	British Journal of Oral & Maxillofacial Surgery	43
8	Applied Sciences-Basel	38
9	Scientific Reports	34
10	Plastic and Reconstructive Surgery	34

AI: Artificial intelligence, OMFS: Oral and maxillofacial surgery

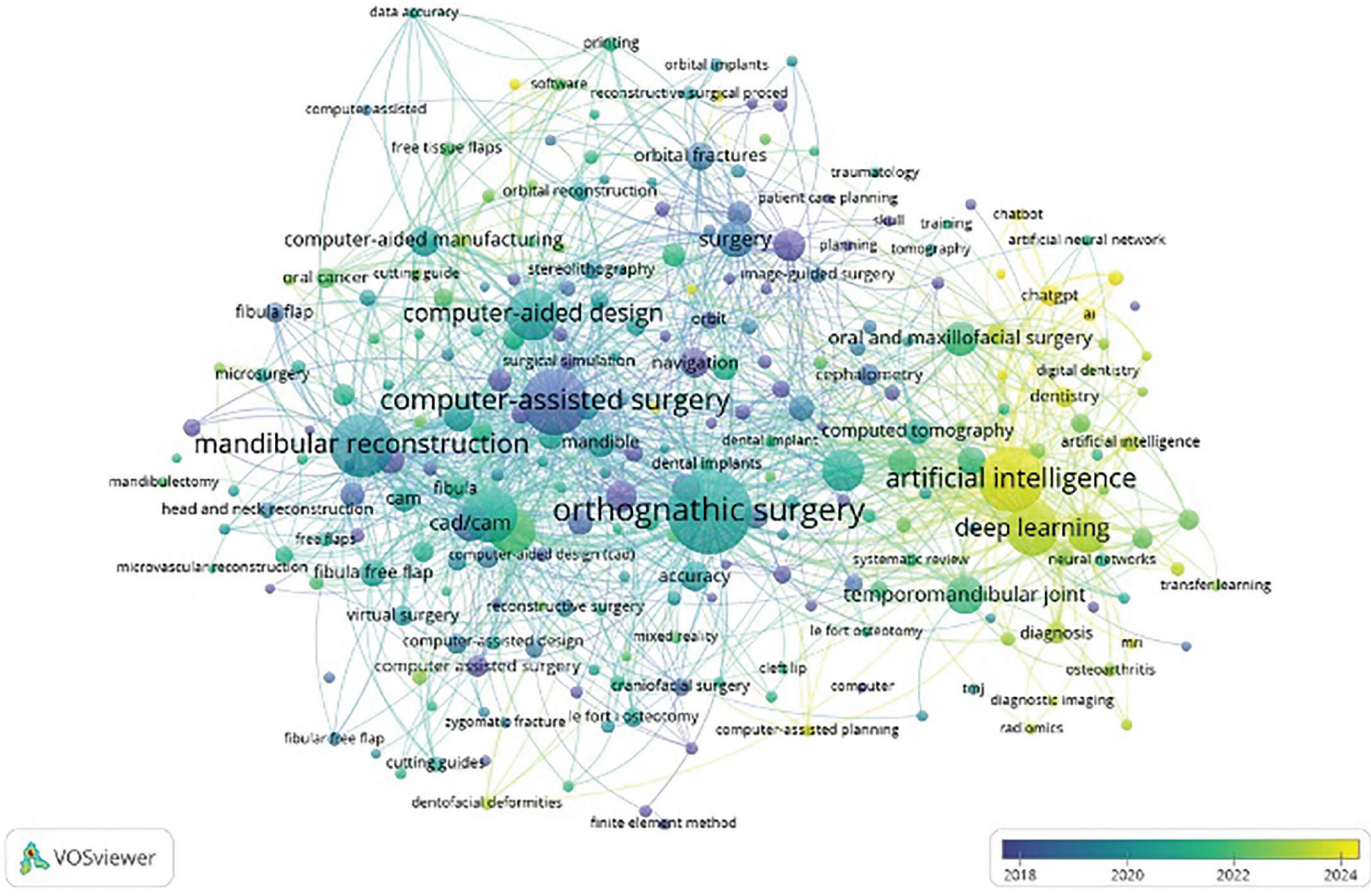


Figure 3. Overlay visualisation of keyword co-occurrence network. Colours indicate the average publication year of studies associated with each keyword, highlighting the shift toward deep learning and virtual surgical planning in recent years

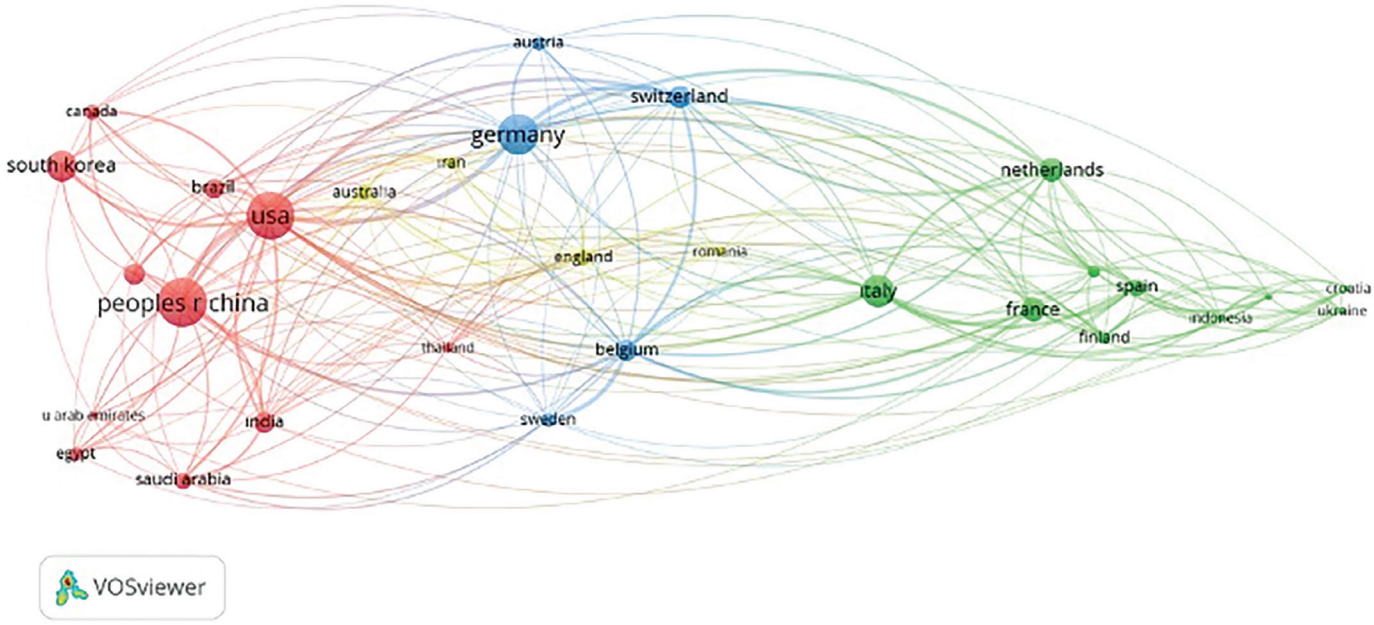


Figure 4. International collaboration network among countries, visualised using VOSviewer. Node size represents publication volume, and link strength reflects the intensity of international co-authorship connections

Table 3. Most frequent keywords related to artificial intelligence applications in oral and maxillofacial surgery

Rank	Keyword	Occurrences
1	Orthognathic surgery	318
2	Computer-assisted surgery	244
3	Computer-aided design/CAD/CAM	224
4	Artificial intelligence	186
5	Mandibular reconstruction	179
6	Virtual surgical planning	172
7	3D printing	126
8	Deep learning	124
9	Maxillofacial surgery	81
10	Surgical navigation	77
11	Machine learning	76
12	Temporomandibular joint	65
13	Accuracy	41
14	Cone-beam computed tomography	37
15	Mandible	36
16	Orbital fractures	36
17	Patient-specific implants	29
18	Facial asymmetry	29
19	Cephalometry	22
20	ChatGPT	20

CAD/CAM: Computer-aided design/computer-aided manufacturing

in previous reviews of AI applications in dentistry and surgical disciplines, where research productivity has been linked to early technological adoption, availability of large digital datasets, and targeted national research funding strategies (1,2,16). The country collaboration network further revealed dense collaborative clusters among high-output countries, suggesting that international co-authorship plays an important role in advancing AI-related OMFS research.

Institutional Productivity and Research Concentration

At the institutional level, AI-related OMFS research was primarily conducted within large academic centres and university-affiliated hospitals. This finding is consistent with observations in the broader healthcare AI literature, where access to advanced computational infrastructure, annotated datasets, and interdisciplinary expertise is often required to develop and validate AI-based systems (18). Systematic reviews focusing on OMFS have similarly noted that methodological innovation and clinical translation are frequently driven by specialised tertiary centres, which may limit the immediate applicability of AI tools in smaller clinical settings (19).

Author Productivity and Collaborative Research Groups

Author productivity analysis indicated that a relatively small group of researchers contributed a substantial proportion of publications in this field. Such authorship patterns are common in emerging research areas and typically reflect the presence of established research groups that function as focal points for

Table 4. The 10 most productive countries in AI-related OMFS publications (2025-2000).

Country	Number of publications	Percentage (%)
People's Republic of China	355	18.16
USA	344	17.60
Germany	252	12.89
Italy	157	8.03
South Korea	144	7.37
Japan	98	5.01
France	92	4.71
Switzerland	85	4.35
Netherlands	82	4.19
Belgium	70	3.58

AI: Artificial intelligence, OMFS: Oral and maxillofacial surgery

Table 5. Leading Institutions in AI-related OMFS Publications (2000-2025)

Rank	Institution	Publications	% of Total records
1	Shanghai Jiao Tong University	87	4.450%
2	Peking University	59	3.018%
3	Sichuan University	50	2.558%
4	Chang Gung Memorial Hospital	49	2.506%
5	Chang Gung University	44	2.251%
6	KU Leuven	40	2.046%
7	Seoul National University	38	1.944%
8	University of Hong Kong	37	1.893%
9	Harvard University	35	1.790%
10	University Hospital Leuven	35	1.790%

AI: Artificial intelligence, OMFS: Oral and maxillofacial surgery

Table 6. Most productive authors in AI-related OMFS publications (2000-2025)

Rank	Author	Publications	% of Total records
1	Gellrich NC	25	1.279%
2	Lo LJ	24	1.228%
3	Marchetti C	24	1.228%
4	Hölzle F	23	1.176%
5	Luo E	23	1.176%
6	Politis C	23	1.176%
7	Modabber A	22	1.125%
8	Schramm A	21	1.074%
9	Sun Y	21	1.074%
10	Lin HH	20	1.023%

AI: Artificial intelligence, OMFS: Oral and maxillofacial surgery

innovation and collaboration (19). Following manual author name disambiguation, no single author was found to dominate the literature overwhelmingly, suggesting that research on AI in OMFS is distributed across multiple active groups rather than centred around a single leading contributor. This distribution may facilitate methodological diversity and promote balanced scientific development within the field.

Thematic Evolution and Emerging Research Trends

Keyword co-occurrence and overlay analyses demonstrated a clear temporal shift in research focus over the study period. Early publications predominantly addressed computer-assisted surgery and basic image-processing techniques, whereas more recent studies increasingly emphasise deep learning, virtual surgical planning, and data-driven decision-support systems. Although not all 3D printing technologies are inherently AI-driven, their frequent integration within AI-supported digital surgical workflows justifies their inclusion in the present analysis. Comparable thematic transitions have been reported in reviews examining AI applications in implant dentistry and related digital workflows (20). Although several studies employed imaging-based data, these approaches were primarily integrated into surgical workflows rather than used as stand-alone radiological analyses, reinforcing the operative orientation of AI research in OMFS. The appearance of generative AI-related keywords, such as ChatGPT, reflects emerging interest in large language models (LLMs) for education, documentation, and decision-support tasks within OMFS. From a bibliometric perspective, the relatively low frequency of generative AI-related keywords compared with established surgical AI themes suggests that LLM-based applications remain an emerging and underrepresented research area within the OMFS literature. While AI has been extensively discussed in dentistry and broader clinical research with respect to its opportunities, limitations, and ethical challenges, the reflection of LLM-focused applications in OMFS publications remains limited. This discrepancy highlights a thematic gap between rapidly evolving AI methodologies and their current representation in OMFS-focused scientific literature, suggesting a potential direction for future research (16). From the perspective of AI research, current bibliometric analysis highlights several key trends and shortcomings. While deep learning-based approaches dominate recent literature, methodological heterogeneity, limited external validation, and reliance on retrospective single-center datasets remain prevalent. Mapping these patterns is essential to guide future AI model development efforts toward standardized evaluation frameworks, transparent reporting, and clinically robust validation strategies. Furthermore, the concentration of research activity in a limited number of

institutions underscores the need for broader interdisciplinary and multicenter collaboration to improve the generalizability of AI-based systems in OMFS (18,21).

Clinical Implications and Future Directions

From a clinical perspective, the observed trends suggest that AI is becoming increasingly embedded in routine OMFS workflows, particularly in orthognathic surgery, mandibular reconstruction, trauma management, and computer-assisted surgical planning. Despite this progress, existing literature remains dominated by retrospective designs and single-centre datasets, which may constrain external validity and real-world implementation (19). Future research efforts should therefore prioritise multicentre collaboration, standardised evaluation protocols, and prospective validation studies to support the safe and effective integration of AI technologies into everyday OMFS practice. From a clinical medicine perspective, the rapid growth of AI-supported OMFS research underscores the increasing integration of digital technologies into surgical decision-making, planning accuracy, and intraoperative guidance. Understanding these trends is essential for clinicians to critically evaluate emerging AI tools and support evidence-based adoption in daily practice.

AI applications supporting imaging-based components of OMFS workflows are dominated by deep learning-based segmentation and classification tasks. Layered convolutional networks have achieved expert-level performance for 3D segmentation of the mandible and other craniofacial structures on CBCT (10,11), and several comparative studies report clinically acceptable accuracy for mandibular canal or defect segmentation, with substantial reductions in manual workload (11,12,14,22).

Taken together, the existing literature suggests that AI has substantial potential to enhance diagnosis, planning precision, and intraoperative guidance in OMFS, but translation into routine clinical practice will depend on rigorous validation, standardisation of workflows, and careful consideration of ethical issues such as data privacy and algorithmic bias (1-4,15,21).

Study Limitations

A limitation of this study is that the bibliometric analysis was based solely on the WoS Core Collection, which may not capture all relevant publications indexed in other databases such as Scopus or PubMed. However, WoS was chosen for its standardized indexing criteria and widespread use in citation-based and science-mapping analyses.

Conclusion

Over the last 25 years, research at the interface of AI and OMFS has expanded rapidly, shifting from early work on stereolithography and navigation to deep learning-based image analysis and fully digital surgical workflows. The present bibliometric analysis shows that orthognathic surgery, mandibular reconstruction, virtual surgical planning, and computer-assisted surgery constitute the main thematic hubs, with deep learning and 3D printing emerging as key enabling technologies.

Although the existing literature demonstrates that AI has considerable potential to enhance diagnostic accuracy, planning precision, and workflow efficiency in OMFS, the evidence base remains dominated by retrospective, single-centre studies. To translate AI tools into routine clinical practice, future research should focus on multicentre prospective validation, rigorous reporting standards, and careful evaluation of clinical and patient-reported outcomes. The science-mapping results presented here can help researchers and clinicians identify influential contributions, recognise gaps in current knowledge, and prioritise high-impact avenues for future investigation.

Notably, the limited representation of ethics-related keywords in the bibliometric findings stands in contrast to the growing ethical complexity of AI applications in OMFS, including data confidentiality, algorithmic bias, and clinical accountability. This mismatch highlights a critical research gap and underscores the need for more explicit and systematic ethical discourse in future OMFS-focused AI research.

Ethics

Ethics Committee Approval: This study followed a bibliometric and science-mapping design and was conducted in accordance with reporting recommendations for bibliometric analyses.

Informed Consent: Only published literature was analysed and no individual patient data were used, ethical approval was not required.

Footnotes

Authorship Contributions

Concept: CG.T., M.K.H., M.Ü., N.A.G., O.Y., Ö.O.G., Design: CG.T., M.K.H., Data Collection or Processing: CG.T., M.K.H., N.A.G., O.Y., Ö.O.G., Analysis or Interpretation: CG.T., M.Ü., N.A.G., O.Y., Ö.O.G., Literature Search: CG.T., M.K.H., N.A.G., O.Y., Ö.O.G., Writing: CG.T., M.K.H., M.Ü.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Dang RR, Kadaikal B, Abbadi SE, Brar BR, Sethi A, Chigurupati R. The current landscape of artificial intelligence in oral and maxillofacial surgery- a narrative review. Oral Maxillofac Surg. 2025;29:37.
2. Rasteau S, Ernenwein D, Savoldelli C, Bouletreau P. Artificial intelligence for oral and maxillo-facial surgery: a narrative review. J Stomatol Oral Maxillofac Surg. 2022;123:276-82.
3. Miragall MF, Knoedler S, Kauke-Navarro M, Saadoun R, Grabenhorst A, Grill FD, et al. Face the future-artificial ıntelligence in oral and maxillofacial surgery. J Clin Med. 2023;12:6843.
4. Abdali HM, Ayoub AY, Omer RA, Alanazi RA, Alsaedi BS, Almutairi LB, et al. Applications of artificial ıntelligence in oral and maxillofacial cosmetic surgery: a systematic review of diagnostic, planning, and outcome assessment tools. Cureus. 2025;17:e92185.
5. Kwon TG. Considerations for virtual surgical planning and simulation in orthognathic surgery. Front Oral Maxillofac Med. 2020;2:15.
6. Alkaabi S, Maningky M, Helder MN, Alsabri G. Virtual and traditional surgical planning in orthognathic surgery - systematic review and meta-analysis. Br J Oral Maxillofac Surg. 2022;60:1184-91.
7. Pampín Martínez MM, Gutiérrez Venturini A, Guiñales Díaz de Cevallos J, Barajas Blanco M, Aragón Niño I, Moreiras Sánchez A, et al. Evaluation of the predictability and accuracy of orthognathic surgery in the era of virtual surgical planning. Applied Sciences. 2022;12:4305.
8. Lee YJ, Oh JH, Kim SG. Virtual surgical plan with custom surgical guide for orthognathic surgery: systematic review and meta-analysis. Maxillofac Plast Reconstr Surg. 2024;46:39.
9. Shalabi MM, Darwich KMA, Kheshfeh MN, Hajeer MY. Accuracy of 3D Virtual surgical planning compared to the traditional two-dimensional method in orthognathic surgery: a literature review. Cureus. 2024;16:e73477.
10. Verhelst PJ, Smolders A, Beznik T, Meewis J, Vandemeulebroucke A, Shaheen E, et al. Layered deep learning for automatic mandibular segmentation in cone-beam computed tomography. J Dent. 2021;114:103786.
11. Ileşan RR, Beyer M, Kunz C, Thieringer FM. Comparison of artificial ıntelligence-based applications for mandible segmentation: from established platforms to ın-house-developed software. Bioengineering. 2023;10:604.
12. Issa J, Konwinska MD, Kazimierczak N, Olszewski R. Assessing the accuracy of artificial intelligence in mandibular canal segmentation compared to semi-automatic segmentation on cone-beam computed tomography images. Pol J Radiol. 2025;90:e172-9.
13. Abesi F, Jamali AS, Zamani M. Accuracy of artificial intelligence in the detection and segmentation of oral and maxillofacial structures using cone-beam computed tomography images: a systematic review and meta-analysis. Pol J Radiol. 2023;88:e256-63.
14. Fan W, Zhang J, Wang N, Li J, Hu L. The application of deep learning on CBCT in dentistry. Diagnostics. 2023;13:2056.
15. Ritto FG, Schmitt ARM, Pimentel T, Canellas JV, Medeiros PJ. Comparison of the accuracy of maxillary position between conventional model surgery and virtual surgical planning. Int J Oral Maxillofac Surg. 2018;47:160-6.
16. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: Chances and Challenges. J Dent Res. 2020;99:769-74.
17. Chen YW, Stanley K, Att W. Artificial intelligence in dentistry: current applications and future perspectives. Quintessence Int. 2020;51:248-57.
18. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med. 2019;25:24-9.
19. Mladenovic R, Pereira L, Pereira A. Artificial Intelligence in oral and maxillofacial surgery: a systematic review. Int. J. Oral Maxillofac. Surg. 2022;51:1-11.
20. Revilla-León M, Gómez-Polo M, Vyas S, Barmak BA, Galluci GO, Att W, et al. Artificial intelligence applications in implant dentistry: a systematic review. J Prosthet Dent. 2023;129:293-300.
21. Taghanaki SA, Abhishek K, Cohen JP, Cohen-Adad J, Hamarneh G. Deep semantic segmentation of natural and medical images: a review. Artif Intell Rev. 2021;54:137-78.
22. Kim SG. The new face of oral and maxillofacial surgery: artificial intelligence's expanding role. J Korean Assoc Oral Maxillofac Surg. 2025;51:127-30.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.40374
Eurasian J Emerg Med. 2026;25: 166-70

Pseudoephedrine and Acute Coronary Events: A Real-World Assessment in Acute Myocardial Infarction Patients

Cemil Yıldız[1], Muhammed İkbal Şaşmaz[1], Mustafa Uçar[2], Özlem Bilir[3], Akkan Avcı[4]

[1]Manisa Celal Bayar University Faculty of Medicine Department of Emergency Medicine, Manisa, Türkiye
[2]Manisa Celal Bayar University Faculty of Medicine, Department of Cardiology, Manisa, Türkiye
[3]Recep Tayyip Erdoğan University Faculty of Medicine, Department of Emergency Medicine, Rize, Türkiye
[4]University of Health Sciences Türkiye, Adana City Training and Research Hospital, Clinic of Emergency Medicine, Adana, Türkiye

Abstract

Aim: The decongestants are frequently prescribed for symptomatic relief to reduce mucosal congestion. However, even in the absence of overt cardiovascular symptoms, patients may subsequently present with serious acute cardiac events. The present study aims to assess the potential association between recent pseudoephedrine exposure and the occurrence of coronary vasospasm in patients presenting with acute myocardial infarction (AMI) to the emergency department, who reported the use of pseudoephedrine-containing products within the preceding week.

Materials and Methods: The study population included patients who presented with chest pain and were diagnosed with AMI [either ST-elevation myocardial infarction (STEMI) or non-STEMI]. The primary objective was to evaluate the history of pseudoephedrine use within the week preceding symptom onset.

Results: Among patients with a history of pseudoephedrine use, the 1-month incidence of major adverse cardiac events (MACE) was 13% (n=3), compared to 12.7% (n=21) in those without such a history. When comparing age, diagnosis, and MACE rates between patients with and without pseudoephedrine use, no statistically significant differences were observed. Regarding MACE subtypes, the most frequent event was death, occurring in 7.4% (n=14) of all patients. Heart failure was identified in 2.6% (n=5), while recurrent myocardial infarction was observed in 2.1% (n=4) patients.

Conclusion: Our findings suggest a clinically relevant association between recent pseudoephedrine use and acute cardiac events in vulnerable patients. This calls for increased awareness among clinicians, pharmacists, and the general public regarding the possible adverse outcomes associated with pseudoephedrine, even when used short-term or at therapeutic doses.

Keywords: Pseudoephedrine, major adverse cardiac events, acute myocardial infarction, percutaneous coronary intervention, emergency

Introduction

Over-the-counter (OTC) decongestants are commonly used worldwide for the symptomatic relief of upper respiratory tract infections, particularly to reduce mucosal congestion. These agents exert their effects through direct or indirect sympathomimetic activity, acting as agonists at both α- and β-adrenergic receptors. When administered orally (e.g., pseudoephedrine, phenylephrine) or intranasally (e.g., oxymetazoline, naphazoline, phenylephrine, ephedrine), they induce vasoconstriction of the nasal mucosa and reduce mucosal edema (1). However, their pharmacologic actions are not confined to the respiratory tract; systemic cardiovascular effects may also occur.

Cardiovascular adverse events, such as hypertension, tachyarrhythmias, and coronary vasospasms, have been documented, particularly in individuals with underlying ischemic heart disease, prompting regulatory restrictions on their use in several countries (2). Among these agents, pseudoephedrine has been particularly associated with serious complications, including

Corresponding Author: Muhammed İkbal Şaşmaz MD, Manisa Celal Bayar University Faculty of Medicine Department of Emergency Medicine, Manisa, Türkiye
E-mail: ikbalsasmaz84@gmail.com **ORCID ID:** orcid.org/0000-0002-3267-3184

Cite this article as: Yıldız C, İkbal Şaşmaz M, Uçar M, Bilir Ö, Avcı A. Pseudoephedrine and acute coronary events: a real-world assessment in acute myocardial infarction patients. Eurasian J Emerg Med. 2026;25: 166-70.

Received: 01.12.2025
Accepted: 13.01.2026
Published: 03.02.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

coronary vasospasm and cerebrovascular events, such as stroke (3-5). These effects are pharmacodynamically consistent with the sympathomimetic profile of the drug. Nevertheless, no clear consensus has not been established, and such adverse outcomes have been reported regardless of patient age, dosage, or vascular health status (4-7). It is also important to consider that pre-existing endothelial dysfunction may serve as a predisposing factor for these events.

In routine clinical practice, especially in outpatient settings, these decongestants are frequently prescribed for symptomatic relief. However, even in the absence of overt cardiovascular symptoms, patients may subsequently present with serious acute cardiac events. Given these risks, it is imperative to re-evaluate the safety profile of pseudoephedrine-containing decongestants, particularly in patients with potential cardiovascular vulnerability.

The present study aims to explore the relationship between pseudoephedrine exposure and reversible coronary artery constriction that may contribute to myocardial ischemia in patients presenting to the emergency department (ED) with acute myocardial infarction (AMI) who reported using pseudoephedrine-containing products within the preceding week.

Materials and Methods

Study Design and Ethical Approval

This observational study was conducted from January 1 to June 30, 2024, in the ED of a tertiary-care hospital. The study population included patients who presented with chest pain and were diagnosed with AMI [either ST-elevation myocardial infarction (STEMI) or non-STEMI (NSTEMI)]. The primary objective was to evaluate the history of pseudoephedrine use within the week preceding symptom onset. Approval was obtained from Manisa Celal Bayar University Ethics Committee prior to study initiation (approval number: E-20478486-050.04-756900, date: 03.04.2024). The study was conducted in accordance with the principles outlined in the Declaration of Helsinki.

Study Population and Variables

During the study period, patients presenting to the ED with chest pain were triaged, monitored, and evaluated by electrocardiography (ECG). Necessary diagnostic tests and initial treatments were administered, and patients with suspected acute coronary syndrome were referred to the cardiology department. Among these, 189 patients were diagnosed with STEMI or NSTEMI and subsequently evaluated for a self-reported history of pseudoephedrine use within the previous seven days. Only patients aged 18 years or older who were conscious, cooperative, oriented, and willing to provide informed consent were included. Patients with myocardial injury secondary to conditions such as hemorrhagic shock or sepsis were excluded from the study.

Demographic data (age, sex, comorbidities), presenting complaints, ECG findings, vital signs at admission, history of pseudoephedrine use, emergency outcomes, cardiology department interventions, and the occurrence of major adverse cardiac events (MACE) within one month among pseudoephedrine-exposed individuals were systematically recorded using a standardized data collection form. MACE was defined as recurrent myocardial infarction, heart failure, ischemic stroke, or death.

Statistical Analysis

All statistical analyses were conducted using the Statistical Package for the Social Sciences, version 26.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics were employed to summarize the data, including frequencies and percentages for categorical variables and the mean, standard deviation (SD), median, and range (minimum-maximum) for continuous variables. The Kolmogorov-Smirnov test was used to assess the normality of continuous variables. For univariate analysis, normally distributed continuous variables were expressed as mean $\pm$ SD and compared using the independent samples t-test. Categorical variables were analyzed using the Pearson's chi-square test. A p value of less than 0.05 was considered indicative of statistical significance.

Results

A total of 189 patients diagnosed with AMI who presented to the ED with chest pain between January and June 2024 were included in the data analysis. The demographic characteristics, comorbidities, clinical presentations, and outcomes of all patients are summarized in Table 1.

The mean age of the study population was 60.88$\pm$13.82 years; the majority of participants were male (n=143, 75.7%). Among these patients, 57.1% (n=108) were diagnosed with NSTEMI. When patients' comorbidities were evaluated, the most common comorbidity was hypertension (53.4%), followed by ischemic heart disease (38.1%) and diabetes mellitus (34.9%). On admission, non-specific ECG changes were the most frequent finding, observed in 45% of cases. Of the 189 patients, 23 (12.2%) reported a history of pseudoephedrine use within the week preceding the onset of AMI symptoms. Among these 23 patients, 8 (34.8%) had a history of ischemic heart disease, 7 (30.4%) hypertension, and 7 (30.4%) diabetes mellitus (7).

Regarding ED outcomes, 87.8% of patients were admitted to the coronary intensive care unit for further monitoring and treatment. One patient died in the ED, while two patients refused further treatment and were discharged against medical advice.

Table 1. Demographic and clinical data of the patients

Age, years, mean ± SD	60.88±13.82	
Gender	Male Female	143 (75.7) 46 (24.3)
Comorbidities	Hypertension Ischemic heart disease Diabetes mellitus Heart failure COPD CKF Stroke Obesity	101 (53.4) 72 (38.1) 66 (34.9) 20 (10.6) 14 (7.4) 11 (5.8) 3 (1.6) 45 (23.8)
Major complaint	Chest pain Shortness of breath Epigastric pain Nausea-vomiting Palpitations Sweating Back pain Numbness in arms Weakness-malaise Syncope	147 (77.8) 15 (7.9) 10 (5.3) 5 (2.6) 3 (1.6) 3 (1.6) 2 (1.1) 2 (1.1) 1 (0.5) 1 (0.5)
Diagnosis	STEMI NSTEMI	81 (42.9) 108 (57.1)
Electrocardiography findings	ST segment elevation ST segment depression T wave inversion Non-specific	80 (42.3) 14 (7.4) 10 (5.3) 85 (45)
History of pseudoephedrine use	Yes No	23 (12.2) 166 (87.8)
Percutaneous coronary intervention	Yes No	171 (90.5) 17 (9)
Intervention type	Balloon-stent Medical follow-up By-pass	110 (58.2) 37 (19.6) 24 (12.7)
One-month MACE	Yes No	24 (12.7) 165 (87.3)

COPD: Chronic obstructive pulmonary disease, CKF: Chronic kidney disease, STEMI: ST-segment elevation myocardial ınfarction, NSTEMI: Non-ST-segment elevation myocardial ınfarction, MACE: Major adverse cardiovascular events, Data were presented as n (%) except age

Following diagnosis, 90.5% of patients (n=171) underwent percutaneous coronary intervention, and 58.2% (n=110) received balloon angioplasty and/or stent placement.

Among patients with a history of pseudoephedrine use, the one-month incidence of MACE was 13% (n=3), compared with 12.7% (n=21) among those without such a history. No statistically significant differences in age, diagnoses, and MACE rates were observed between patients with and without pseudoephedrine use (Table 2).

Regarding MACE subtypes, the most frequent event was death, occurring in 7.4% (n=14) of all patients. Heart failure was identified in 2.6% (n=5) of patients, while recurrent myocardial infarction was observed in 2.1% (n=4) of patients.

Discussion

This study provides valuable insights into the potential cardiovascular risks of pseudoephedrine use, especially in populations with underlying vulnerabilities. Pseudoephedrine, a sympathomimetic agent commonly included in OTC cold and allergy medications, is frequently perceived as benign due to its widespread availability and symptomatic efficacy. However, our findings highlight that a non-negligible proportion (12.2%) of patients presenting with AMI had used pseudoephedrine within the preceding week, underscoring the importance of reevaluating the cardiovascular safety of pseudoephedrine in susceptible individuals.

Although no statistically significant association was identified between pseudoephedrine use and the incidence of coronary vasospasm in AMI patients, this observation must be interpreted with caution due to the limited sample size. Importantly, a considerable proportion of patients exposed to pseudoephedrine had pre-existing cardiovascular comorbidities: 34.8% had ischemic heart disease, 30.4% had hypertension, and 30.4% had diabetes mellitus. These conditions are well-known contributors to endothelial dysfunction, which plays a critical role in the pathogenesis of vasospasm and thrombosis, particularly in the presence of vasoconstrictive agents such as pseudoephedrine (8,9). Although definitive causality cannot be determined from

Table 2. Comparison of patients with and without a history of pseudoephedrine use

History of pseudoephedrine use	**Yes (n=23)**	**No (n=166)**	**p value**
Age, years, mean ± SD	58.52±18.04	61.20±13.17	0.384
Diagnosis (n, %) STEMI NSTEMI	 8 (34.8) 15 (65.2)	 73 (44) 93 (56)	0.404
1-month MACE, n (%) Pozitive Negative	 3 (13) 20 (87)	 21 (12.7%) 145 (87.3%)	0.958

SD: Standard deviation STEMI: ST-segment elevation myocardial ınfarction, NSTEMI: Non-ST-segment elevation myocardial ınfarction, MACE: Major adverse cardiovascular events

our cross-sectional design, the data suggest that pseudoephedrine may precipitate ischemic events in predisposed individuals.

One of the most compelling aspects of our findings is the continued use of pseudoephedrine despite advanced age and significant cardiovascular comorbidities patient profiles for which regulatory bodies such as the Food and Drug Administration and European Medicines Agency have issued cautionary guidelines (10). This points to a concerning gap in clinical awareness and risk stratification regarding the systemic effects of pseudoephedrine, particularly in emergency and primary care settings where comprehensive medication histories may not always be obtained. Easy access to pseudoephedrine-containing products exacerbates this risk, particularly among individuals who self-medicate without consulting healthcare professionals.

While establishing a causal link between pseudoephedrine use and AMI through long-term prospective studies is methodologically complex, the real-world data presented here provide a compelling signal. In particular, our findings align with prior literature and case reports suggesting that pseudoephedrine can induce coronary vasospasm, arrhythmias, and cerebrovascular events via sympathomimetic mechanisms (11,12). In our study, pseudoephedrine users frequently exhibited non-specific electrocardiographic changes, such as ST-segment alterations, which may reflect transient myocardial ischemia or vasospastic phenomena. Although we did not evaluate QT intervals or utilize ambulatory electrocardiographic monitoring, this absence underscores a broader gap in the assessment of the cardiovascular effects of OTC medications in clinical practice and highlights the need for more comprehensive monitoring in future studies.

Furthermore, the 1-month rate of MACE was comparable between pseudoephedrine users (13%) and non-users (12.7%). However, the presence of serious outcomes including death among recently exposed individuals is clinically significant and warrants attention. These findings support the hypothesis that pseudoephedrine may not act as a direct etiological factor but rather as a physiological stressor capable of triggering cardiovascular events in patients with preexisting endothelial vulnerability or atherosclerotic disease.

In addition to cardiovascular concerns, pseudoephedrine's potential neurotoxicity must be considered. Although our study primarily focused on cardiac outcomes, it is worth noting that the mean age of the cohort was 60 years. Elderly individuals are particularly vulnerable to the systemic effects of sympathomimetics. Pseudoephedrine has been associated with central nervous system depression, seizures, and even fatal outcomes, especially in patients with pre-existing neurological conditions (7,13). A notable case report describes an 83-year-old woman who developed prolonged seizures after ingesting 120 mg of pseudoephedrine; her symptoms persisted for three months (14). These findings underscore the need for heightened vigilance when prescribing or recommending OTC medications in geriatric populations.

Study Limitations

Our study has several limitations, most notably its single-center design and relatively small sample size, which may limit generalizability. In addition, the absence of data on patients' baseline vital signs, vascular health, renal and hepatic function, as well as detailed information on pseudoephedrine dosage, frequency, and route of administration introduces further uncertainty. Furthermore, the relatively small number of patients reporting pseudoephedrine use and the absence of detailed dosage information limit the strength of the conclusions. Additionally, given the lack of statistically significant differences, the findings should be interpreted with caution and not be considered evidence of a causal relationship.

Conclusion

Pseudoephedrine remains a widely used agent for symptomatic relief in conditions such as the common cold and allergic rhinitis. However, its systemic sympathomimetic effects particularly in elderly individuals and those with cardiovascular or neurological comorbidities carry potential risks that are frequently underrecognized. Although not statistically significant, our findings suggest a clinically relevant association between recent pseudoephedrine use and acute cardiac events among vulnerable patients. This calls for increased awareness among clinicians, pharmacists, and the general public of the possible adverse outcomes associated with pseudoephedrine, even when used in the short term or at therapeutic doses. Patient education, enhanced pharmacovigilance, and more stringent regulatory oversight may be warranted to mitigate these risks. Particularly in high-risk populations, medication histories should be carefully evaluated before initiating even seemingly benign symptomatic therapies.

Ethics

Ethics Committee Approval: Approval was obtained from Manisa Celal Bayar University Ethics Committee prior to study initiation (approval number: E-20478486-050.04-756900, date: 03.04.2024). The study was conducted in accordance with the principles outlined in the Declaration of Helsinki.

Informed Consent: Only patients aged 18 years or older who were conscious, cooperative, oriented, and willing to provide informed consent were included.

Footnotes

Authorship Contributions

Surgical and Medical Practices: C.Y., M.İ.Ş., Concept: C.Y., M.İ.Ş., A.A., Design: M.İ.Ş., M.U., Data Collection or Processing: C.Y., Analysis or Interpretation: C.Y., M.İ.Ş., Ö.B., A.A., Literature Search: C.Y., Ö.B., A.A., Writing: C.Y., M.İ.Ş., M.U.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

Data Availability Statement

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request

References

1. Erickson CH, McLeod RL, Mingo GG, Egan RW, Pedersen OF, Hey JA. Comparative oral and topikal decongestant effects of phenylpropanolamine and pseudoephedrine. Am J Rhinol. 2001;15:83-90.
2. Grimaldi-Bensouda L, Begaud B, Benichou J, Nordon C, Dialla O, Morisot N, et al. Decongestant use and the risk of myocardial infarction and stroke: a case – crossover study. Sci Rep. 2021;11:4160.
3. Lafaurie M, Olivier P, Khouri C, Atzenhoffer M, Bihan K, Durrieu G, et al. Myocardial infarction and ischemic stroke with vasoconstrictors used as nasal decongestant for common cold: a French pharmacovigilance survey. Eur J Clin Pharmacol. 2020;76:603-4.
4. Manini AF, Kabrhel C, Thomsen TW. Acute myocardial infarction after over-the-counter use of pseudoephedrine. Ann Emerg Med. 2005;45:213-6.
5. Cantu C, Arauz A, Murillo-Bonilla LM, López M, Barinagarrementeria F. Stroke associated with sympathomimetics contained in over the counter cough and cold drugs. Stroke. 2003;34:1667-72.
6. Browning MG, Seddon JO, Yung LTM, Gough G. An unusual case of systemic cardiovascular side effects from the application of over-the-counter nasal congestion drops. BMJ Case Rep. 2011;2011:bcr0620103078.
7. Glowacka K, Wiela-Hojenska A. Pseudoephedrine-benefits and risk. Int J Mol Sci. 2021;22:5146.
8. Virani SS, Newby LK, Arnold SV, Bittner V, Brewer LC, Demeter SH, et al. 2023 AHA/ACC/ACCP/ASPC/NLA/PCNA guideline for the management of patients with chronic coronary disease: a report of the American Heart Association/American College of Cardiology Joint Committee on Clinical Practice Guidelines. Circulation. 2023;148:e9-e119. Epub 2023 Jul 20. Erratum in: Circulation. 2023;148:e148. Erratum in: Circulation. 2023;148:e186.
9. Tarkin JM, Dweck MR, Evans NR, Takx RA, Brown AJ, Tawakol A. Imaging atherosclerosis. Circ Res. 2016;118:750-69.
10. FDA, U. (2013, January). FDA issues public health warning on phenylpropanolamine. Avaible from: https://www.fda.gov/drugs/information-drug-class/fda-issues-public-health-warning-phenylpropanolamine
11. Arama V, Ganea O, Neagu D, Rosculet C, Arama SS. Update on the efficiency and safety of orally administered nasal decongestants. Ro J Infect Dis. 2023;26:125-34.
12. Głowacka, K, Wiela-Hojeńska A. (2021). Pseudoephedrine—benefits and risks. International Journal of Molecular Sciences. 2021;22:5146.
13. Douhard R, Humbert P, Milon JY, Pegahi R. Pharmacology of chlorphenamine and pseudoephedrine use in the common cold: a narrative review. Curr Med Res Opin. 2024;40:2125-37.
14. Mudarres F. Behind the counter danger: Pseudoephedrine-induced seizure in an elderly patient. Cureus. 2024;16:e67099.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Case Report
DOI: 10.4274/eajem.galenos. 2026.81904
Eurasian J Emerg Med. 2026;25: 171-4

Hypertensive Encephalopathy in the Emergency Department, a Case of Posterior Reversible Encephalopathy Syndrome (PRES)

Jeremy Zhenwen Pong, Aliviya Dutta, Jeevan Raaj Thangayah

Singapore General Hospital, Department of Emergency Medicine, Singapore

Abstract

Hypertension is commonly encountered in the emergency department. While emergency physicians are understandably concerned about hypertensive emergencies, such occurrences are rare. Here, we present a case report of hypertensive encephalopathy causing posterior reversible encephalopathy syndrome (PRES), which manifested in a patient as altered mental status and seizures. We also discuss risk factors for PRES, common presentations, investigations, management, and PRES as a crucial consideration in patients with hypertension and altered mental status, given differences in management compared with other differential diagnoses. With timely recognition and treatment, most patients with PRES recover completely within several days.

Keywords: Hypertension, hypertensive encephalopathy, hypertensive emergency, posterior reversible encephalopathy syndrome, emergency department

Introduction

Hypertension is commonly encountered in the emergency department (ED), either as a presenting complaint or as an incidental finding at triage. A hypertensive emergency is defined as a potentially life-threatening condition with end-organ damage secondary to markedly elevated blood pressure (1). This commonly occurs when blood pressure rises above 180/120 mmHg, causing end-organ injury in the ophthalmic, renovascular, cardiovascular, or cerebrovascular systems (2). Such patients may present with visual deficits, chest pain, shortness of breath, limb numbness or weakness, altered mental status, seizures, or even coma. Although emergency physicians are understandably concerned about hypertensive emergencies, such occurrences are rare and were found to account for less than 2% of all hypertensive cases in the ED (3,4). Hypertensive encephalopathy causing posterior reversible encephalopathy syndrome (PRES) is even rarer, poisted to result from hypertension-induced autoregulatory failure leading to cerebral edema that preferentially affects the parietal and occipital lobes (5). Distinguishing PRES from other diagnoses (e.g. ischemic stroke) is crucial because subsequent management differs and can significantly affect patient outcomes. Here, we present a case of PRES manifesting as altered mental status and seizures.

Case Report

Ms A was a 65-year-old woman with a history of diabetes mellitus, hypertension, and hyperlipidemia. While vacationing aboard a cruise ship, she failed to disembark when the ship docked. Ms A was subsequently found semiconscious in her cabin with fecal incontinence. Assessment by the onboard cruise doctor noted that she was drowsy, had a blood pressure of 227/105 mmHg, and a capillary blood glucose level of 27 mmol/L (486 mg/dL). She was transferred via ambulance to the hospital.

On arrival at the ED, Ms A was afebrile and hypertensive, with a blood pressure of 169/87 mmHg, a heart rate of 82, a respiratory rate of 18, and an oxygen saturation of 96% on room air. Though drowsy, she provided a brief history of feeling unwell the day before, although she was unable to specify exact symptoms. Notably, she admitted to omitting her chronic medications for diabetes and hypertension for one week (metformin 1000 mg twice daily and telmisartan 40 mg once daily in the morning).

Corresponding Author: Jeremy Zhenwen Pong MD, Singapore General Hospital, Department of Emergency Medicine, Singapore
E-mail: jeremypongzw@gmail.com **ORCID ID:** orcid.org/0000-0001-6011-0584

Cite this article as: Pong JZ, Dutta A, Thangayah JR. Hypertensive encephalopathy in the emergency department, a case of posterior reversible encephalopathy syndrome (PRES). Eurasian J Emerg Med. 2026;25: 171-4.

Received: 24.12.2025
Accepted: 19.01.2026
Published: 06.02.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

On physical examination, Ms A was drowsy but rousable, with a Glasgow coma scale (GCS) score of 13 (E3 V4 M6). Fecal incontinence was noted. Examination of the heart, lungs, and abdomen were unremarkable, and no overt neurological deficits or injuries were found.

Laboratory investigations showed an elevated serum glucose concentration of 25.1 mmol/L (452 mg/dL) (reference range: 3.9-11 mmol/L), with serum ketones marginally elevated at 0.8 mmol/L (reference range: 0-0.6 mmol/L), and venous blood gas pH and bicarbonate normal at 7.35 (reference range: 7.35-7.45) and 23 mmol/L (reference range: 19-29 mmol/L) respectively. There were mildly elevated transaminases (aspartate aminotransferase 76 U/L, reference range 12-42 U/L; alanine transaminase 194 U/L, reference range 6-66 U/L), markedly elevated alkaline phosphatase (1729 U/L, reference range 39-99 U/L), and normal bilirubin (21 µmol/L, reference range 7-32 µmol/L). The remainder of the full blood count and the renal panel with extended electrolytes were otherwise unremarkable. Bedside ultrasound examinations of the heart and abdomen were grossly normal. A computed tomography scan of the brain demonstrated small hypodensities in the right frontal lobe that were of indeterminate age which may have represented chronic lacunar infarcts. While still in the ED, Ms A became tremulous and developed asterixis. She subsequently experienced a witnessed generalised tonic-clonic seizure, which was swiftly aborted with 5 mg of intravenous diazepam. Systolic blood pressure for the patient was by now trending consistently above 220 mmHg.

Clinical impression was that of altered mental status possibly secondary to PRES, hyperlgycemia not in crisis, and deranged liver enzymes for investigation. An intravenous nicardipine infusion was started for blood pressure control, as was insulin therapy for the hyperglycemia, and levetiracetam for seizure control. Consequently, the patient appeared more alert as blood pressure began to decline, though she remained slightly confused with a GCS of 14 (E4, V4, M6).

In view of the patient's persistent drowsiness, seizure, and elevated blood pressure requiring intravenous antihypertensives, the Neurology department was informed with a view toward an early inpatient magnetic resonance imaging (MRI) of the brain. An MRI brain scan performed the following day showed findings suggestive of PRES: areas of cortical and subcortical white matter signal changes with associated swelling, predominantly in the posterior distribution, involving both occipital lobes (Figure 1a). Though Ms A's usual antihypertensive medication was initially held off on admission, she was restarted on them after the MRI diagnosis of PRES was confirmed. Ms A underwent a repeat brain MRI approximately two weeks after the initial MRI scan, with repeat MRI demonstrating near-complete resolution of the previously noted hyperintense areas (Figure 1b). She also made a full clinical recovery.

Discussion

PRES was named for the radiological and clinical features of white matter edema in the posterior cerebrum on brain MRI, and its potential for neurological symptoms to be reversed should prompt and appropriate treatment be initiated (6). The mechanism by which PRES and cerebral swelling result in

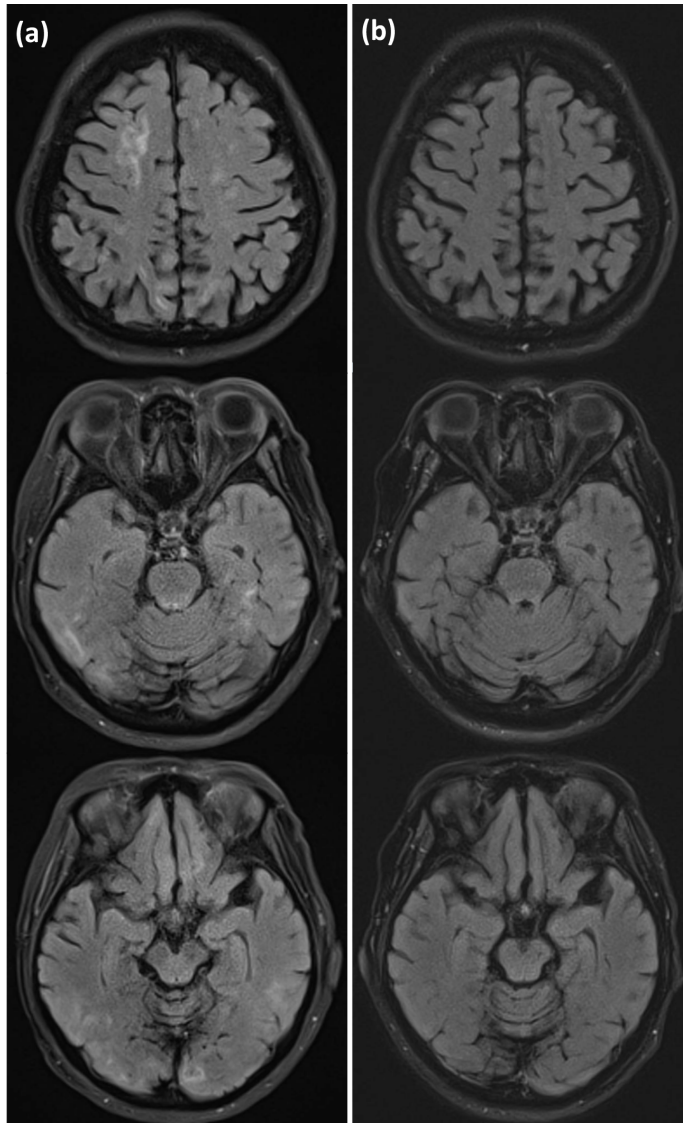


Figure 1. (a) Magnetic resonance imaging (MRI) of the brain on patient presentation, showing areas of T2W/FLAIR hyperintensity in the subcortical white matter and cortex predominantly in a posterior distribution involving both occipital lobes, which may suggest posterior reversible encephalopathy syndrome. Similar signal changes in the right temporal and bilateral frontal lobes. (b) Near resolution of T2W/FLAIR hyperintense areas on repeat MRI two weeks later.

neurological dysfunction is not well understood, but is postulated to occur when excessive hypertension damages the endothelial lining of the blood-brain barrier, leading to cerebral edema. This posterior predilection may reflect reduced autoregulatory capacity, supporting the "vasogenic" theory of PRES (7). Though PRES commonly occurs as a complication of hypertension, it has also been reported in normotensive patients who are immunosuppressed (e.g. transplant recipients), in patients with sepsis, in those with impaired renal function, in autoimmune conditions, and in association with blood transfusion and drugs of abuse. In these instances, cerebral swelling is postulated to result from cytotoxicity or autoimmune attack on cerebral blood vessels (6,8). This represents the "toxic theory" of PRES.

Recognition of PRES as the cause of neurological disturbance in the ED requires a high index of suspicion because the differential diagnoses for patients presenting with altered mental status are broad, and initial history and investigations may be unrevealing. This is important, as the management of PRES differs from that of ischemic stroke, another common cause of altered mental status in the ED. In the former, we aim to reduce blood pressure, whereas in the latter, permissive hypertension is allowed. This presents a conundrum, and the clinical features of each case are therefore paramount in determining the most likely etiology of the patient's altered mental status.

What then are the clinical predictors of PRES? In a retrospective cohort study of 220 patients presenting with acute neurological symptoms, statistically significant associations were found between PRES and symptoms such as epileptic seizures, encephalopathy, visual disturbances, as well as with risk factors such as hypertension, chemotherapy, and renal failure (9). A model using the above clinical parameters for prediction of PRES achieved an accuracy (area under the curve) of 0.793 (9). Patients with medical conditions such as systemic lupus erythematosus with renal involvement, glomerulonephritis, and patients receiving immunosuppressive treatment for transplantation or rheumatologic conditions were also at higher risk for developing PRES. A review article on the prevalence of symptoms and signs in patients with PRES noted that seizures (60-75%), encephalopathy (50-80%), headache (50%), visual disturbances (33%), focal neurological deficits (10-15%), and status epilepticus (5-15%) were the most common clinical manifestations (10). As the symptoms and signs of PRES are non-specific and can also be observed in a variety of other disorders, clinical judgment is critical for rapid identification and management. An expedited MRI of the brain should be strongly considered when the clinical suspicion of PRES is high.

Treatment of PRES revolves around the identification and removal of precipitating factors where possible (chemotherapy drugs, immunosuppressants), management of hypertension with titratable intravenous antihypertensives such as nicardipine or labetalol, aiming for a gradual reduction in blood pressure of approximately 25% from the presenting value over the first few hours, and antiepileptic medications for seizures if present (10). In studies examining the prognosis of patients with PRES, 75% to 90% of patients made a full recovery, with most recovering within two to eight days (10,11).

Conclusion

Hypertension and altered mental status are common presentations to the ED. The diagnosis of PRES or hypertensive encephalopathy requires a high index of suspicion and should be considered in patients with risk factors such as hypertension, chemotherapy administration, or renal failure, presenting with common symptoms and signs of PRES, such as seizures or status epilepticus, encephalopathy, headache, visual disturbances, or focal neurological deficits. In such patients, a brain MRI should be performed as it most accurately depicts disease-specific neuroradiological findings and helps to guide subsequent management. With timely recognition and treatment, most patients with PRES exhibit a complete recovery in several days.

Ethic

Informed Consent: This is a retrospective study.

Footnotes

Author Contributions

Surgical and Medical Practices: J.Z.P., A.D., J.R.T., Concept: J.Z.P., A.D., J.R.T., Design: J.Z.P., A.D., J.R.T., Data Collection or Processing: J.Z.P., A.D., J.R.T., Analysis or Interpretation: J.Z.P., A.D., J.R.T., Literature Search: J.Z.P., Writing: J.Z.P., A.D., J.R.T.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Potter T, Schaefer TJ. Hypertensive encephalopathy. StatPearls. Treasure Island FL: 2023, StatPearls Publishing LLC. 2023.
2. Miller J, McNaughton C, Joyce K, Binz S, Levy P. Hypertension management in emergency departments. Am J Hypertens. 2020;33:927-34.
3. Johnson W, Nguyen ML, Patel R. Hypertension crisis in the emergency department. Cardiol Clin. 2012;30:533-43.
4. Suneja M, Sanders ML. Hypertensive emergency. Med Clin North Am. 2017;101:465-78. Epub 2017 Mar 2.
5. Tetsuka S, Ogawa T. Posterior reversible encephalopathy syndrome: a review with emphasis on neuroimaging characteristics. J Neurol Sci. 2019;404:72-9. Epub 2019 Jul 17.

6. Zelaya JE, Al-Khoury L. Posterior reversible encephalopathy syndrome. 2024. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan–.
7. Fugate JE, Claassen DO, Cloft HJ, Kallmes DF, Kozak OS, Rabinstein AA. Posterior reversible encephalopathy syndrome: associated clinical and radiologic findings. Mayo Clin Proc. 2010;85:427-32.
8. Triplett JD, Kutlubaev MA, Kermode AG, Hardy T. Posterior reversible encephalopathy syndrome (PRES): diagnosis and management. Pract Neurol. 2022;22:183-9.
9. Faille LD, Fieuws S, Van Paesschen W. Clinical predictors and differential diagnosis of posterior reversible encephalopathy syndrome. Acta Neurol Belg. 2017;117:469-75.
10. Fugate JE, Rabinstein AA. Posterior reversible encephalopathy syndrome: clinical and radiological manifestations, pathophysiology, and outstanding questions. Lancet Neurol. 2015;14:914-25. Epub 2015 Jul 13. Erratum in: Lancet Neurol. 2015;14:874.
11. Liman TG, Bohner G, Endres M, Siebert E. Discharge status and in-hospital mortality in posterior reversible encephalopathy syndrome. Acta Neurol Scand. 2014;130:34-9. Epub 2013 Dec 13.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.12244
Eurasian J Emerg Med. 2026;25: 175-85

Metabolic Augmentation of mREMS: Improved 30-day Mortality Prediction through Lactate and Base Excess Integration in Emergency Patients

Gürbüz Meral[1], Serkan Günay[2], Ramazan Özel[3], Ahmet Köse[4]

[1]Bulancak State Hospital, Clinic of Emergency Medicine, Giresun, Türkiye
[2]Hitit University Çorum Erol Olçok Training and Research Hospital, Clinic of Emergency Medicine, Çorum, Türkiye
[3]Yıldırım Beyazıt University Yenimahalle Training and Research Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[4]Rize State Hospital, Clinic of Emergency Medicine, Rize, Türkiye

Abstract

Aim: To evaluate whether adding base excess (BE), alone or in combination with lactate, to the modified Rapid Emergency Medicine score (mREMS) improves 30-day mortality prediction in emergency department (ED) red-zone patients.

Materials and Methods: This observational cohort study included eligible ED red-zone patients presenting between May 2018 and December 2019 (n=1798). Vital signs, Glasgow Coma scale (GCS) scores, and blood gas parameters were collected. Because the relationship between BE and 30-day mortality was non-linear, BE was categorized using a data-driven, mortality-associated classification approach. The resulting categories were weighted using effect sizes from multivariable logistic regression and converted into an ordinal variable (0-3 points), which was integrated into the mREMS framework. Three models were developed: mREMS-Be, mREMS-L, and mREMS-LBe. Model discrimination was evaluated using the area under the curve (AUC) and comparisons were performed using the DeLong test.

Results: Thirty-day mortality was associated with older age, lower systolic blood pressure, lower SpO_2, and lower GCS ($p<0.05$). BE was not predictive when modeled as a continuous variable; however, it demonstrated an independent association with mortality when modeled as an ordinal variable ($p=0.022$). The AUC of the baseline mREMS was 0.753; it increased to 0.767 with the addition of BE and to 0.782 when BE was combined with lactate. Integration of BE into mREMS resulted in a significant improvement in model discrimination in both the overall and the non-traumatic red-zone patient groups.

Conclusion: Incorporating BE into mREMS provides complementary prognostic value by improving 30-day mortality discrimination among high-acuity ED patients.

Keywords: mREMS, lactate, base excess, mortality prediction, emergency department, risk stratification

Introduction

Early recognition of clinical deterioration in emergency departments (EDs) is critically important for ensuring patient safety and for directing advanced therapeutic resources efficiently (1). Accordingly, Early Warning scores (EWS), which rely on physiological parameters, are widely used as complementary tools to support the rapid and systematic identification of high-risk patients (2). However, it is well established that systems based solely on vital signs may be limited, particularly during the early phases of metabolic derangement when clinical manifestations have not yet emerged (3). For this reason, efforts to integrate metabolic biomarkers into physiologic scoring systems have gained increasing importance with the aim of improving diagnostic accuracy (4).

In this context, blood gas lactate is a powerful metabolic biomarker reflecting tissue hypoperfusion and anaerobic metabolism, and it has been shown to be closely associated with mortality in sepsis, trauma, and various critical illnesses (4,5). Base excess (BE) in

Corresponding Author: Gürbüz Meral MD, Bulancak State Hospital, Clinic of Emergency Medicine, Giresun, Türkiye
E-mail: gurbuzmeral61@gmail.com **ORCID ID:** orcid.org/0000-0002-8026-4609
Cite this article as: Meral G, Günay S, Özel R, Köse A. Metabolic augmentation of mREMS: improved 30-day mortality prediction through lactate and base excess integration in emergency patients. Eurasian J Emerg Med. 2026;25: 175-85.

Received: 03.12.2025
Accepted: 23.01.2026
Published: 23.02.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

blood gas analysis, on the other hand, reflects the severity of metabolic acidosis and serves as a comprehensive indicator of metabolic stress, particularly in patients with comorbidities or active compensatory mechanisms (6-8).

Among EWS models, the modified Rapid Emergency Medicine score (mREMS) incorporates practical parameters such as age, vital signs and neurological status and is frequently used to predict mortality in the ED (9). Although the current literature does not definitively establish mREMS as the optimal score for ED populations, its component variables and predictive potential make it a useful model (9,10). While it was initially validated in trauma patients, efforts have been made to broaden its use to general ED cohorts (9-11). Recently, Donoso Calero et al. (12) demonstrated that integrating lactate into mREMS improved its predictive performance in a general ED patient population. Considering the prognostic value of BE, its representation of a metabolic dimension distinct from that of lactate, and the theoretical advantage of combining these biomarkers to achieve a more comprehensive risk assessment, BE emerges as an additional biomarker with potential for integration into EWS models such as mREMS (13-15).

This study evaluates BE-augmented versions of mREMS in a large cohort of red-zone patients from a tertiary ED using a three-level triage system (red-yellow-green) (16). Building on established evidence for lactate integration, the primary objective is to assess the incremental prognostic value of BE, alone and in combination with lactate, for 30-day mortality. Secondary objectives include comparison with the original mREMS and between BE-augmented models. The findings are expected to support earlier identification of high-risk ED patients, more effective use of limited healthcare resources, and inform the development of metabolically augmented mREMS variants that may be suitable for future integration into digital and automated clinical decision-support systems.

Materials and Methods

Study Design and Data Source

This observational cohort study evaluated the 30-day mortality risk among adult ED patients classified as red-zone cases. The study population consisted of eligible red-zone patients who presented to a tertiary training and research hospital in Turkey between 10 May 2018 and 12 December 2019 and for whom all predefined inclusion criteria were met and prospectively recorded study parameters were available. A previous publication from our group utilized only a subset of this dataset comprising 868 patients who met ESI 1-2 criteria; (17), however, the present study includes all eligible red-zone patients who fulfilled the predefined inclusion criteria and had complete data available for analysis (n=1798), and differs from the earlier work in terms of sampling frame, analytical strategy, and study objectives. Accordingly, the current study represents a retrospective secondary analysis of prospectively recorded data.

The study complied with the Declaration of Helsinki. Written informed consent had been obtained during the initial prospective data collection. For the current retrospective analysis of the entire cohort, additional approval was obtained from the Scientific Research Ethics Committee of the University of Health Sciences Türkiye, Trabzon Faculty of Medicine (decision no: 2024/11, date: 05.11.2024), and renewed consent was not required. All data were anonymized prior to analysis and handled in accordance with institutional data confidentiality policies.

Study Population

Patients aged ≥18 years who were classified as red-zone cases in the ED, who had complete vital signs recorded at initial presentation, who had available arterial or venous blood gas analyses were included. Completeness was defined as the availability of all variables required to calculate mREMS at initial presentation. Exclusion criteria were: presentation with cardiac arrest, pregnancy, missing mandatory parameters for mREMS calculation, and duplicate visits. The patient inclusion and exclusion steps, with corresponding numbers, are shown in Figure 1.

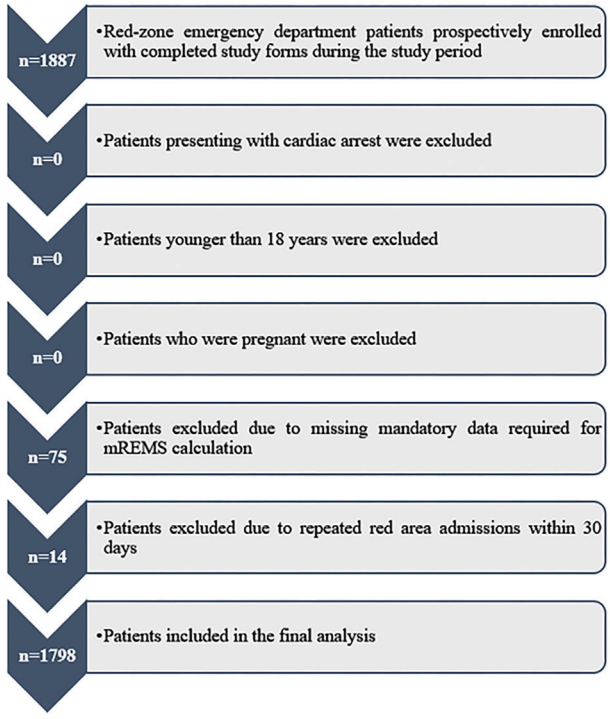


Figure 1. Flow Chart
mREMS: Modified Rapid Emergency Medicine score

Vital parameters were measured by triage nurses, and Glasgow Coma scale (GCS) was assessed by an emergency physician. Blood gas analyses were performed using a Radiometer ABL90 FLEX device integrated within the ED. Thirty-day mortality was verified through the national Death Notification System. Etiological classification (infectious or traumatic) was independently assessed by three emergency medicine specialists, and final categorization was established based on agreement between at least two specialists.

BE Classification and Scoring Method

Given that the association between BE and mortality has been shown in the literature and in logistic regression analyses of the present dataset to be non-linear, the inclusion of BE in the model as a continuous variable was considered inappropriate (8,18,19). Therefore, a data-driven approach was adopted to examine how the relationship between BE distribution and mortality risk varied across BE ranges.

Accordingly, BE values in the overall cohort were stratified into quintiles based on their distribution, and 30-day mortality rates were calculated for each quintile. This exploratory assessment indicated a non-linear risk distribution, with higher mortality observed at both markedly negative and markedly positive BE values. Adjacent quintiles exhibiting similar mortality rates were subsequently merged, resulting in four BE categories representing progressively increasing levels of risk.

In weighting the derived BE categories, reliance on crude mortality rates alone was avoided, and an approach accounting for the effects of potential confounders was employed. To this end, BE categories were entered into a multivariable logistic regression model together with established clinical variables, including age, vital signs, and GCS score. The adjusted odds coefficients obtained from this model were used to guide the assignment of ordinal risk points to the BE categories. Based on the observed risk gradient, BE categories were scored from 0 to 3 as follows:

- BE 2.2 to 4.6 mmol/L: 0 point
- BE ≥0.1 to <2.2 mmol/L or >4.6 mmol/L: 1 point
- BE ≥-3.2 to <0.1 mmol/L: 2 points
- BE <-3.2 mmol/L: 3 points

This approach ensured that BE classification and weighting were grounded in the risk structure observed within the study cohort and adjusted for relevant covariates, rather than relying on externally predefined thresholds. However, acknowledging that the prognostic signal of BE may be sensitive to the modeling approach, the maximum BE score was capped at 3 points to avoid disproportionate weighting that could obscure the prognostic contribution of lactate within the composite score.

Following score assignment, BE was treated as an ordinal variable and re-entered into multivariable logistic regression analysis. In this analysis, ordinal BE demonstrated a statistically significant association with 30-day mortality. This finding suggests that the derived BE categories exhibit internal consistency with respect to risk direction. Consequently, BE was integrated into mREMS and mREMS-based models as a complementary metabolic parameter reflecting a non-linear risk pattern.

Statistical Analysis

All primary statistical analyses were conducted using JAMOVI (v. 2.6.44). Distributional characteristics of continuous variables were assessed using the Shapiro-Wilk test in conjunction with histogram and Q-Q plot inspection. Normally distributed variables were summarized as mean ± standard deviation, whereas non-normally distributed variables were summarized as median (interquartile range). Categorical variables were summarized as counts and percentages. Between-group comparisons were performed using the Mann-Whitney U test for continuous variables and χ^2 test for categorical variables.

To identify independent predictors of 30-day mortality, two multivariable binomial logistic regression models were constructed. In the first model, BE was included as a categorical variable (BE_{0-3}) to estimate category-specific adjusted odds ratios (ORs) relative to the lowest-risk reference group. In the second model, BE was incorporated as an ordinal variable (0-3 points), reflecting a progressively increasing risk of mortality based on adjusted effect sizes observed in the categorical model. This sequential modeling strategy enabled the evaluation of both category-specific effects and monotonic risk behavior for BE, while minimizing assumptions about linearity.

For multivariable models, regression coefficients (β), OR, 95% confidence intervals (CIs), and p values were reported. ORs and corresponding 95% CIs were derived directly from regression coefficients [$OR = \exp(\beta)$, 95% $CI = \exp(\beta \pm 1.96 \times \text{standard error})$]. Model fit was assessed using the Akaike information criterion and McFadden's R^2.

Model discrimination was assessed using ROC curves and area under the curve (AUC) values. Optimal cut-off points for each score were determined using the Youden index to provide an objective balance between sensitivity and specificity. Differences in AUC among the four evaluated scores (mREMS, mREMSBe, mREMS-L, and mREMS-LBe) were analyzed using pairwise DeLong tests. A p value<0.05 was considered statistically significant for all analyses.

To address potential overfitting associated with data-driven BE categorization, internal validation of all mREMS-based models was performed using bootstrap resampling with 5.000 iterations. Calibration performance, particularly for BE-containing models, was further evaluated using decile-based calibration plots in the overall study population.

Results

A total of 1,798 red-zone ED patients were included in the study (Figure 1). The most frequent diagnoses in the cohort were pneumonia (23.0%), cerebrovascular disease (14.5%), and multiple trauma (13.7%) (Table 1). The overall population was predominantly elderly, and non-survivors were significantly older than survivors (76 vs. 67 years, p<0.001; Table 2). Although the sex distribution was comparable between groups, respiratory rate and heart rate were higher in the mortality group, whereas systolic blood pressure (SBP), mean arterial pressure (MAP), and SpO_2 were lower (all p<0.001). GCS was also significantly lower among non-survivors (Table 2).

Evaluation of blood gas parameters revealed pronounced metabolic impairment in patients who died. BE values were more negative, and although median lactate levels were identical between survivors and non-survivors, non-survivors demonstrated a wider interquartile range and higher upper lactate values (both p<0.001; Table 2).

Distribution of mREMS and its metabolically enhanced variants showed that all scores were significantly higher among non-survivors (all p<0.001; Table 2). Greater score separation between survivors and non-survivors was observed for mREMS-L and mREMS-LBe compared with the baseline mREMS.

In multivariable logistic regression analysis, age, SBP, SpO_2, and GCS were identified as independent predictors of 30-day mortality [all p<0.05; Table 3 (a-b)]. BE when entered as a continuous variable was not independently associated with 30-day mortality and therefore did not demonstrate a significant linear effect. However, when modeled as an ordinal variable, BE showed an independent association with mortality (OR per category increase: 1.200; p=0.022; Table 3b). Lactate remained a strong independent predictor across all models (p<0.001).

In the ROC analysis, the baseline mREMS model demonstrated discriminative performance, yielding an AUC of 0.753. Augmentation of mREMS with BE resulted in a modest improvement in discrimination (AUC=0.767), while integration

Table 1. Most frequent clinical diagnoses according to ICD classification

ICD	Count	% of total
Pneumonia	414	23.0%
Cerebrovascular disease	261	14.5%
Multiple trauma	237	13.7%
Acute abdomen	167	9.3%
Gastrointestinal bleeding	66	3.7%
Acute renal failure	57	3.2%
Complicated urinary tract infection	49	2.7%
Non-STEMI	48	2.7%
Sepsis	44	2.4%
Intoxication	40	2.2%
Pulmonary edema	33	1.8%
COPD with acute exacerbation	31	1.7%
STEMI	29	1.6%
Dyspnea	29	1.6%
Spontaneous pneumothorax	27	1.5%
Acute arterial occlusion	25	1.4%
Acute pancreatitis	25	1.4%
Arrhythmia	22	1.2%
Pulmonary embolism	20	1.1%
Other (<1%)	174	9.3%

ICD: International Statistical Classification of Diseases and Related Health Problems, Non-STEMI: Non-ST-segment elevation myocardial infarction, STEMI: ST-segment elevation myocardial infarction, COPD: Chronic obstructive pulmonary disease

Table 2. Baseline characteristics of the study population according to 30-day mortality				
Variable	**Overall cohort (n=1798)**	**Survivors (n=1560)**	**Non-survivors (n=238)**	**p value**
Demographics				
Age, years	73±20	67±20	76±18	**<0.001**
Male sex, n (%)	965 (53.6)	845 (54.2)	120 (50.4)	0.313
Vital signs				
Temperature, °C	37 (36-37)	37 (36-37)	37 (36-37)	0.441
Heart rate, bpm	86 (75-100)	86 (75-100)	90 (75-110)	**0.022**
Respiratory rate/min	18 (15-22)	18 (15-21)	20 (15-28)	**<0.001**
SpO_2, %	95 (90-98)	95 (92-98)	92 (85-96)	**<0.001**
SBP, mmHg	130 (110-150)	130 (110-150)	110 (90-134)	**<0.001**
MAP, mmHg	93 (80-106)	93 (83-106)	80 (70-96)	**<0.001**
Neurological				
GCS	15 (15-15)	15 (15-15)	14 (9-15)	**<0.001**
Blood gas parameters				
Base excess, mmol/L	1 (-2-4)	1 (-2-4)	-1 (-6-3)	**<0.001**
Lactate, mmol/L	2 (1-2)	2 (1-2)	2 (1-4)	**<0.001**
Clinical scores				
mREMS	4 (3-7)	4 (2-6)	7 (5-11)	**<0.001**
mREMS-Be	5 (3-7)	5 (3-7)	8 (5-12)	**<0.001**
mREMS-L	6 (4-9)	6 (4-8)	10 (7-15)	**<0.001**
mREMS-LBe	7 (5-9)	6 (4-9)	11 (7-15)	**<0.001**
Clinical course				
Intubation on arrival, n (%)	36 (2.0)	16 (1.0)	20 (8.4)	**<0.001**
Hospitalization, n (%)	1672 (93.0)	1421 (91.2)	235 (98.7)	**<0.001**
ICU admission, n (%)	489 (27.2)	344 (22.1)	145 (60.9)	**<0.001**
Etiology				
Infective, n (%)	757 (42.1)	651 (41.7)	106 (44.5)	0.455
Traumatic, n (%)	247 (13.7)	227 (14.6)	20 (8.4)	**0.014**

SBP: Systolic blood pressure, MAP: Mean arterial pressure, GCS: Glasgow Coma scale, mREMS: Modified Rapid Emergency Medicine score, mREMS-Be: Modified REMS with base excess, mREMS-L: Modified REMS with lactate, mREMS-LBe: Modified REMS with lactate and base excess, ICU: Intensive care unit

Table 3a. Multivariable binomial logistic regression analysis for 30-day mortality with categorical base excess							
Predictor	**Estimate**	**SE**	**Z**	**p**	**OR**	**95% CI**	
						Lower	**Upper**
Intercept	3.22766	1.21392	2.659	0.008	25.220	2.336	272.299
Age	0.02583	0.00518	4.987	<0.001	1.026	1.016	1.037
HR	-0.00505	0.00370	-1.365	0.172	0.995	0.988	1.002
RR	0.01402	0.01132	1.238	0.216	1.014	0.992	1.037
SpO_2	-0.03781	0.00974	-3.881	<0.001	0.963	0.945	0.981
SBP	-0.01297	0.00261	-4.973	<0.001	0.987	0.982	0.992
GCS	-0.17668	0.02276	-7.763	<0.001	0.838	0.801	0.876
BE-categoric							
BE_1-BE_0	0.09246	0.24780	0.373	0.709	1.097	0.675	1.783
BE_2-BE_0	0.42258	0.26623	1.587	0.112	1.526	0.906	2.571
BE_3-BE_0	0.47844	0.26375	1.814	0.070	1.614	0.962	2.706
Lactate	0.15577	0.03662	4.254	<0.001	1.169	1.088	1.256

BE: Base excess, SE: Standard error, OR: Odds ratio, CI: Confidence interval, SBP: Systolic blood pressure, HR: Heart rate, RR: Respiratory rate, GCS: Glasgow Coma scale

Table 3b. Multivariable binomial logistic regression analysis for 30-day mortality with ordinal base excess

Predictor	Estimate	SE	Z	p	OR	95% CI	
						Lower	Upper
Intercept	3.17117	1.19578	2.65	0.008	23.835	2.288	248.356
Age	0.02576	0.00518	4.98	<0.001	1.026	1.016	1.037
HR	-0.00509	0.00370	-1.38	0.169	0.995	0.988	1.002
RR	0.01388	0.01132	1.23	0.220	1.014	0.992	1.037
SpO_2	-0.03760	0.00970	-3.88	<0.001	0.963	0.945	0.982
SBP	-0.01296	0.00261	-4.97	<0.001	0.987	0.982	0.992
GCS	-0.17584	0.02272	-7.74	<0.001	0.839	0.802	0.877
BE-ordinal	0.18193	0.07921	2.30	0.022	1.200	1.027	1.401
Lactate	0.15526	0.03627	4.28	<0.001	1.168	1.088	1.254

Two multivariable binomial logistic regression models evaluating the association between base excess (BE) and 30-day mortality are presented. The primary outcome was 30-day all-cause mortality, coded as a binary variable (1= death, 0= survival).
In the first model, BE was included as a categorical variable, derived using data-driven mortality-based thresholds, to estimate category-specific odds ratios. In the categorical model, BE was grouped into four mortality-BE_{0-3} derived from a data-driven threshold analysis. BE_0 represents the reference category associated with the lowest observed mortality risk. BE_1, BE_2, and BE_3 represent progressively higher-risk BE strata based on increasing 30-day mortality rates. These adjusted odds ratios were subsequently used to assign weighted risk scores to each BE category.
In the second model, BE was incorporated as an ordinal variable (0-3 points) reflecting increasing mortality risk based on the magnitude of odds ratios observed in the categorical model. For the ordinal BE variable, the reported odds ratio represents the change in the odds of 30-day mortality associated with a one-category increase in BE score. Mortality was coded as 1 (event) and 0 (non-event). Estimates (β), SEs, Wald Z statistics, ORs, and 95% CIs for ORs are reported. The model was estimated with n=1.798
OR: Odds ratio, CI: Confidence interval, SE: Standard error, HR: Heart rate (beats per minute), RR: Respiratory rate (breaths per minute), SBP: Systolic blood pressure (mmHg), GCS: Glasgow Coma scale, BE: (mmol/L), Lactate (mmol/L)

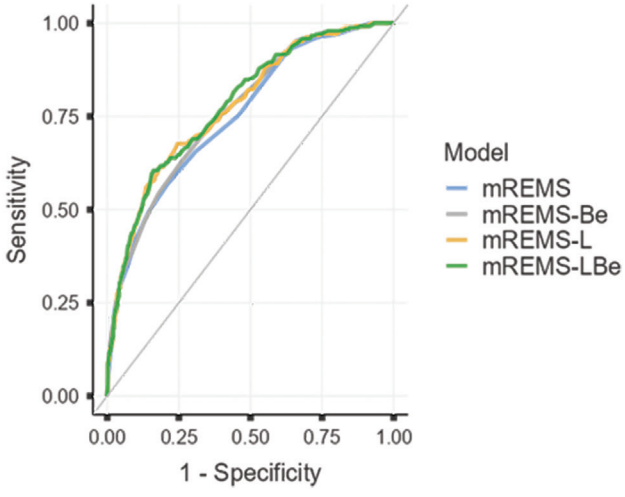


Figure 2a. Combined ROC Curves of mREMS and mREMS-based models in the overall cohort

mREMS: modified Rapid Emergency Medicine score, mREMS-Be: Modified REMS with base excess, mREMS-L: Modified REMS with lactate, mREMS-LBe: Modified REMS with lactate and base excess

of lactate led to a further increase in AUC to 0.778. The combined mREMS-LBe achieved an AUC of 0.782. Although mREMS-LBe yielded the numerically highest AUC, its performance was not statistically superior to that of mREMS-L (Table 4). Pairwise DeLong tests demonstrated that mREMS-Be, mREMS-L, and mREMS-LBe all performed significantly better than mREMS (p=0.006, 0.002, and 0.002, respectively). While mREMS-LBe also outperformed mREMS-Be (p=0.014), it did not show a statistically significant advantage over mREMS-L (p=0.228). ROC curves for the overall cohort are shown in Figure 2a.

Internal validation using bootstrap resampling demonstrated minimal optimism across all models. Optimism-corrected AUCs were 0.7663 for mREMS-Be and 0.7816 for mREMS-LBe, closely approximating apparent performance. Model accuracy assessed by the Brier score was 0.0976 for mREMS-Be and 0.0966 for mREMS-LBe. Decile-based calibration analyses demonstrated good agreement between predicted and observed 30-day mortality across risk strata in the overall cohort (Figure 3a-b; Table 5). In subgroup analyses, all models demonstrated high discriminative ability in trauma patients (AUCs: 0.873 for mREMS, 0.865 for mREMS-Be, 0.846 for mREMS-L, and 0.837 for mREMS-LBe), however, pairwise DeLong test comparisons revealed no statistically significant differences between the models (all p>0.05; Table 4). In the non-traumatic subgroup, integration of both BE and lactate yielded higher AUC values compared with baseline mREMS (p<0.05 for all relevant DeLong comparisons). Although mREMS-LBe showed the numerically highest AUC in this subgroup as well, it did not show a significant advantage over mREMS-L. ROC curves for the traumatic and non-traumatic subgroups are presented in Figures 2b and 2c, respectively.

Overall, metabolically augmented mREMS models demonstrated improved discrimination in predicting 30-day mortality compared with the baseline score.

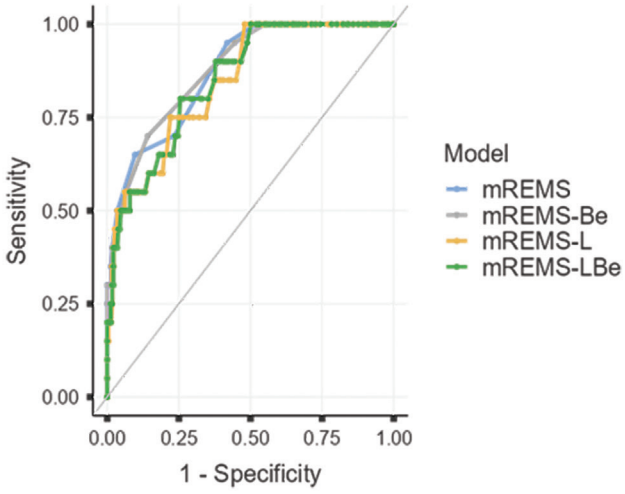


Figure 2b. Combined ROC curves of mREMS and mREMS-based models in the traumatic subgroup

mREMS: modified Rapid Emergency Medicine score, mREMS-Be: Modified REMS with base excess, mREMS-L: Modified REMS with lactate, mREMS-LBe: Modified REMS with lactate and base excess

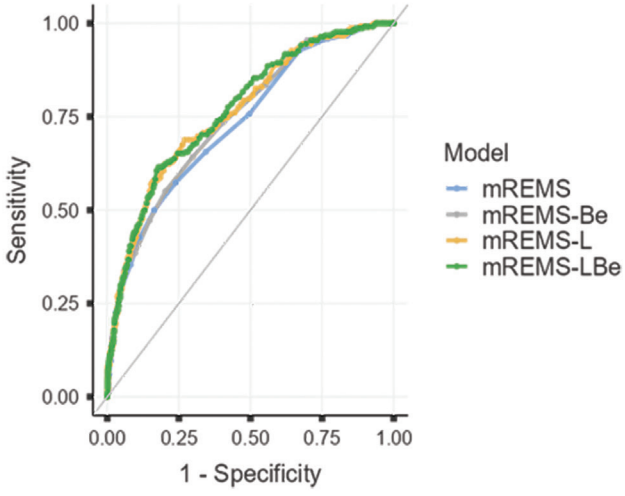


Figure 2c. Combined ROC curves of mREMS and mREMS-based models in the non-traumatic subgroup

mREMS: modified Rapid Emergency Medicine score, mREMS-Be: Modified REMS with base excess, mREMS-L: Modified REMS with lactate, mREMS-LBe: Modified REMS with lactate and base excess

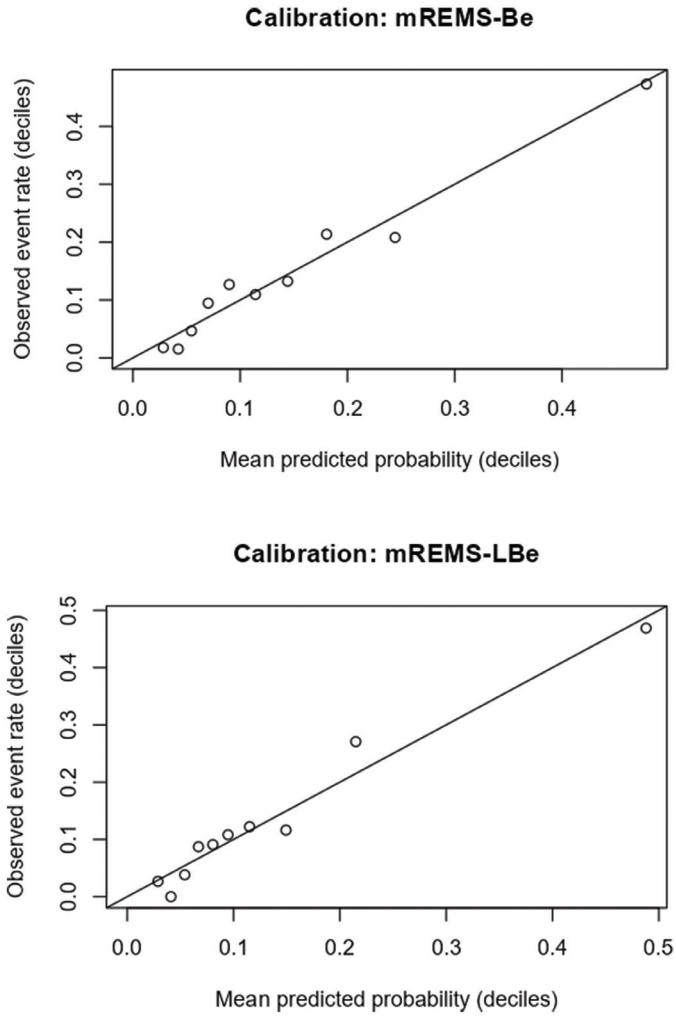


Figure 3 (a-b). Decile-based calibration plots of the mREMS-Be and mREMS-LBe models in the overall cohort

Decile-based calibration plots for the mREMS-Be and mREMS-LBe models in the overall cohort following bootstrap internal validation (B=5.000). Predicted 30-day mortality risks are plotted against observed event rates within deciles of predicted risk. The dashed diagonal line represents ideal calibration.

mREMS: modified Rapid Emergency Medicine score, mREMS-Be: Modified REMS with base excess, mREMS-L: Modified REMS with lactate, mREMS-LBe: Modified REMS with lactate and base excess

Table 4. ROC analysis with DeLong test comparisons of mREMS and mREMS-based models in the overall cohort and clinical subgroups

	Model	AUC	95% CI	Cut-off	Sensitivity	Specificity	Model comparisons with DeLong test			
							vs	mREMS-Be	mREMS-L	mREMS-LBe
Overall cohort	mREMS	0.753	0.720-0.786	≥**7.0**	56.7%	79.0%	mREMS	**0.006***	**0.002***	**0.002***
	mREMS-Be	0.767	0.735-0.799	≥**7.0**	63.4%	73.5%	mREMS-Be	-	0.124*	**0.014***
	mREMS-L	0.778	0.746-0.810	≥**9.0**	61.3%	81.9%	mREMS-L	-	-	0.228*
	mREMS-LBe	0.782	0.750-0.813	≥**9.9**	60.5%	83.9%	mREMS-LBe	-	-	-
Traumatic subgroup	mREMS	0.873	0.800-0.945	≥**6.0**	65.0%	90.3%	mREMS	0.683*	0.083*	0.233*
	mREMS-Be	0.865	0.790-0.940	≥**6.0**	70.0%	85.9%	mREMS-Be	-	0.140*	0.054*
	mREMS-L	0.846	0.765-0.928	≥**6.8**	75.0%	78.0%	mREMS-L	-	-	0.777*
	mREMS-LBe	0.837	0.752-0.923	≥**6.8**	80.0%	74.4%	mREMS-LBe	-	-	-
Non-traumatic subgroup	mREMS	0.733	0.697-0.770	≥**8.0**	50.0%	83.6%	mREMS	**0.007***	**<0.001***	**<0.001***
	mREMS-Be	0.750	0.715-0.785	≥**8.0**	55.0%	79.7%	mREMS-Be	-	**0.028***	**0.002***
	mREMS-L	0.766	0.731-0.800	≥**9.0**	62.4%	79.8%	mREMS-L	-	-	0.236*
	mREMS-LBe	0.772	0.738-0.806	≥**9.9**	61.5%	82.2%	mREMS-LBe	-	-	-

*DeLong p values represent pairwise comparisons of ROC AUCs between the model listed in each row and the corresponding model listed in the column header. P values are reported in the intersecting cells. Optimal cut-off values were determined using the Youden index
AUC: Area under the curve, mREMS: Modified Rapid Emergency Medicine Score, mREMS-Be: Modified REMS with base excess, mREMS-L: Modified REMS with lactate, mREMS-LBe: Modified REMS with lactate and base excess

Table 5. Apparent and optimism-corrected performance metrics of mREMS and BE-augmented mREMS models

	Model	N	Events	AUC (apparent)	AUC (optimism-corrected)	Calibration intercept (corrected)	Calibration slope (corrected)	Brier (apparent)	Brier (corrected)
Overall cohort	mREMS	1798	238	0.7530	0.7530	0.0102	1.004	0.0989	0.0993
	mREMS-Be	1798	238	0.7669	0.7671	0.0102	1.003	0.0976	0.0980
	mREMS-L	1798	238	0.7779	0.7783	0.0088	1.004	0.0969	0.0971
	mREMS-LBe	1798	238	0.7819	0.7819	0.0052	1.002	0.0966	0.0969
Traumatic subgroup	mREMS	247	20	0.8728	0.8731	-0.0120	0.9804	0.0510	0.0529
	mREMS-Be	247	20	0.8649	0.8648	0.0042	0.9871	0.0522	0.0541
	mREMS-L	247	20	0.8461	0.8459	0.0115	0.9895	0.0579	0.0604
	mREMS-LBe	247	20	0.8371	0.8377	0.0119	0.9956	0.0576	0.0595
Non-traumatic subgroup	mREMS	1551	218	0.7335	0.7340	0.0123	1.006	0.1063	0.1066
	mREMS-Be	1551	218	0.7499	0.7501	0.0081	1.004	0.1047	0.1051
	mREMS-L	1551	218	0.7658	0.7655	0.0066	1.001	0.1030	0.1036
	mREMS-LBe	1551	218	0.7716	0.7715	0.0042	1.002	0.1027	0.1030

The Apparent AUC and Brier score represent model performance in the original sample. Optimism-corrected values were obtained using bootstrap internal validation with 5.000 resamples. Calibration intercept and slope values closer to 0 and 1, respectively, indicate better calibration.
AUC: Area under the curve, mREMS, modified Rapid Emergency Medicine score, mREMS-Be, mREMS augmented with base excess, mREMS-L, mREMS augmented with lactate, mREMS-LBe, mREMS augmented with lactate and base excess, N, number of patients, Events, number of deaths within 30 days; CI: Confidence interval, mREMS-Be: Modified REMS with base excess, mREMS-L: Modified REMS with lactate, mREMS-LBe: Modified REMS with lactate and base excess, BE: Base excess

Discussion

This study demonstrates that integrating mREMS-Be significantly improves 30-day mortality prediction among critically ill ED patients. The mREMS model enriched with lactate and BE provides superior discrimination compared with mREMS-Be and demonstrates predictive performance comparable to that of the previously validated mREMS-L model in critically ill ED patients not limited to trauma.

Integration of mREMS-L is not a novel concept. A recent multicenter study by Donoso Calero et al. (12) in a mixed ED population showed that mREMS-L significantly improved the performance of the original mREMS. Our results were consistent with those reported in this study. These findings suggest that mREMS, which has demonstrated strong prognostic capability primarily in trauma populations, may also be applicable to non-trauma ED populations when combined with lactate (9,10,12).

In non-traumatic etiologies, clinical deterioration may progress more insidiously and vital signs may lag behind early metabolic impairment; consequently, classical EWS models such as mREMS may be less sensitive in the early recognition of clinical deterioration (20-22). In this context, lactate, as a marker of metabolic stress and early hypoperfusion (4,23) may help mitigate this limitation. As another clinically informative component of blood gas analysis reflecting metabolic reserve, BE has clear theoretical relevance; however, to our knowledge, its integration into an EWS has not been widely evaluated (7,8,18,19,24). Although mREMS does not represent all EWSs, incorporating BE into mREMS allowed us to explore this question within a pragmatic and commonly used framework. Given lactate's established prognostic capacity and BE's ability to reflect metabolic reserve, the combined use of these biomarkers may offer complementary clinical information.

On the other hand, the question of whether BE or lactate better predicts clinical or metabolic deterioration in critically ill patients remains unresolved (25-28). Caputo et al. (26) reported that lactate and BE were equivalent in predicting shock among trauma patients with normal vital signs, whereas other studies have shown context-dependent differences, with lactate or BE demonstrating relative advantages in specific populations (28,29). Several investigations further suggest that combining lactate and BE may improve the prediction of clinical deterioration compared with either biomarker alone (8,23). Accordingly, lactate and BE were incorporated into the mREMS framework both standalone and in combination.

Our results show that BE, consistent with the literature, did not exhibit a linear association with mortality, but became prognostically informative when modeled as an ordinal variable (4,19,24). This pattern may reflect threshold-dependent depletion of metabolic reserve rather than a uniform linear effect. Accordingly, ordinally scored BE, both alone and combined with lactate, improved prognostic accuracy when integrated into mREMS in ED patients, not limited to trauma. These findings are also consistent with evidence from previous systematic reviews and meta-analyses demonstrating that the integration of metabolic biomarkers into physiology-based scoring systems enhances the prediction of clinical deterioration in critically ill patients, not limited to trauma (20,21,30). In the overall cohort and the non-traumatic subgroup, mREMS-LBe demonstrated superior discriminative performance compared with those of the original mREMS and mREMS-Be. The lack of a significant performance advantage over mREMS-L suggests that the observed improvement is largely attributable to the dominant prognostic contribution of lactate (18,23). In this context, BE appears to function as a complementary metabolic biomarker rather than as an independent determinant of prognostic discrimination, and integration of BE and lactate may represent an option for enhanced risk stratification in general ED populations, pending external validation.

In the trauma cohort, incorporation of BE, lactate, or both biomarkers did not yield a statistically significant improvement in the discriminative performance of mREMS. This finding may reflect limited statistical power due to the relatively small trauma sample size. Nonetheless, prior studies have demonstrated that BE performs comparably to lactate and may be superior in specific clinical contexts for predicting mortality and adverse outcomes in trauma populations (29,31,32). Given the established prognostic utility of mREMS in trauma patients, further evaluation of BE integration in larger, adequately powered trauma cohorts is warranted as part of future model refinement efforts.

In non-trauma populations, relatively high mortality rates and the limited ability of conventional vital sign–based parameters to reliably detect early physiological deterioration highlight the increasing importance of integrating metabolically sensitive biomarkers, such as lactate and BE, into physiology-based EWS (20-22,33). However, in the context of the present study, translating this physiological rationale into a statistically robust and clinically applicable modeling strategy required several methodological assumptions, including data-driven categorization of BE. Although this approach was intended to account for the non-linear association between BE and 30-day mortality, the derivation of cohort-specific thresholds may introduce optimism bias and limit model transportability to other ED populations. While bootstrap-based internal validation and calibration analyses demonstrated minimal optimism

and acceptable internal stability, these methods do not ensure external validity. Therefore, the performance of BE-augmented mREMS models should be interpreted with caution, and multicenter external validation is warranted before broader clinical application.

Study Limitations

Limitations of this study include the relatively small trauma subgroup, which prevented adequate statistical power; the possibility that arterial-venous differences in blood gas sampling introduced additional variability in BE; and the single-center retrospective design, which may limit generalizability. Thus, validation of the integration of metabolic biomarkers into mREMS requires multicenter, prospective studies.

Conclusion

This study demonstrates that integration of BE into mREMS, whether alone or combined with lactate, is associated with improved discrimination in predicting 30-day mortality among high-acuity ED patients, with BE serving as a complementary metabolic parameter rather than supplanting the established prognostic role of lactate. Following external validation, BE-augmented mREMS models may be considered for integration into automated clinical decision-support systems in acute care settings.

Ethics

Ethics Committee Approval: The study protocol was approved by the Scientific Research Ethics Committee of the University of Health Sciences Türkiye, Trabzon Faculty of Medicine (decision no: 2024/11, date: 05.11.2024).

Informed Consent: Written informed consent had been obtained during the initial prospective data collection.

Footnotes

Authorship Contributions

Surgical and Medical Practices: G.M., Concept: G.M., Design: G.M., Data Collection or Processing: G.M., S.G., R.Ö., A.K., Analysis or Interpretation: G.M., S.G., Literature Search: G.M., R.Ö., A.K., Writing: G.M., S.G., R.Ö., A.K.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. O'Neill SM, Clyne B, Bell M, Casey A, Leen B, Smith SM, et al. Why do healthcare professionals fail to escalate as per the early warning system (EWS) protocol? A qualitative evidence synthesis of the barriers and facilitators of escalation. BMC Emerg Med. 2021;21:15.
2. Nielsen PB, Schultz M, Langkjaer CS, Kodal AM, Pedersen NE, Petersen JA, et al. Adjusting early warning score by clinical assessment: a study protocol for a Danish cluster-randomised, multicentre study of an individual early warning score (I-EWS). BMJ Open. 2020;10:e033676.
3. Jo S, Jeong T, Lee JB, Jin Y, Yoon J, Park B. Validation of modified early warning score using serum lactate level in community-acquired pneumonia patients. The national early warning score-lactate score. Am J Emerg Med. 2016;34:536-41. Epub 2015 Dec 23.
4. Schork A, Moll K, Haap M, Riessen R, Wagner R. Course of lactate, pH and base excess for prediction of mortality in medical intensive care patients. PLoS One. 2021;16:e0261564.
5. Moncada-Gutiérrez D, Vásquez-Tirado GA, Meregildo-Rodríguez ED, Quispe-Castañeda CV, Cuadra-Campos M, Abanto-Montalván PH, et al. Lactate-enhanced-qSOFA (LqSOFA) score as a predictor of in-hospital mortality in patients with sepsis: systematic review and meta-analysis. Eur J Trauma Emerg Surg. 2025;51:33.
6. Prasad H, Vempalli N, Agrawal N, Ajun UN, Salam A, Subhra Datta S, et al. Correlation and agreement between arterial and venous blood gas analysis in patients with hypotension-an emergency department-based cross-sectional study. Int J Emerg Med. 2023;16:18.
7. Zander R. Base excess (BE): reloaded. Eur J Med Res. 2024;29:281.
8. Smuszkiewicz P, Jawień N, Szrama J, Lubarska M, Kusza K, Guzik P. Admission lactate concentration, base excess, and alactic base excess predict the 28-day inward mortality in shock patients. J Clin Med. 2022;11:6125.
9. Barbosa GS, Gaspar JO, Nogueira LS, Vieira RCA, Lima KP, Sousa RMC. Performance of the modified rapid emergency medicine score in patients with and without traumatic brain injury. J Neurosci Nurs. 2022;54:130-5. Epub 2022 Mar 4.
10. Phunghassaporn N, Sukhvibul P, Techapongsatorn S, Tansawet A. Accuracy and external validation of the modified rapid emergency medicine score in road traffic injuries in a Bangkok level I trauma center. Heliyon. 2022;8:e12225.
11. Miller RT, Nazir N, McDonald T, Cannon CM. The modified rapid emergency medicine score: a novel trauma triage tool to predict in-hospital mortality. Injury. 2017;48:1870-7. Epub 2017 Apr 25.
12. Donoso Calero MI, Mordillo-Mateos L, Martín-Conty JL, Polonio-López B, López-González Á, Durantez-Fernández C, et al. Modified rapid emergency medicine score-lactate (mREMS-L) performance to screen non-anticipated 30-day-related-mortality in emergency department. Eur J Clin Invest. 2023;53:e13994. Epub 2023 Apr 8.
13. Zhou Q, Miao Y, Li F, Liu J, Li J, Li N. Association between base excess and mortality in septic older adults. Geriatr Gerontol Int. 2025;25:380-6. Epub 2025 Jan 20.
14. Cheng Y, Zhang Y, Tu B, Qin Y, Cheng X, Qi R, et al. Association between base excess and mortality among patients in ICU with acute kidney injury. Front Med (Lausanne). 2021;8:779627.
15. Liu J, Chen Y, Yang B, Zhao J, Tong Q, Yuan Y, et al. Association between alactic base excess on mortality in sepsis patients: a retrospective observational study. J Intensive Care. 2025;13:20.
16. Pasli S, Yadigaroğlu M, Kirimli EN, Beşer MF, Unutmaz İ, Ayhan AÖ, et al. ChatGPT-supported patient triage with voice commands in the emergency department: a prospective multicenter study. Am J Emerg Med. 2025;94:63-70. Epub 2025 Apr 17.

17. Meral G, Ardıç Ş, Günay S, Güzel K, Köse A, Durmuş HG, et al. Comparative analysis of glasgow coma scale, quick sepsis-related organ failure assessment, base excess, and lactate for mortality prediction in critically ill emergency department patients. Turk J Emerg Med. 2024;24:231-7.
18. Qi J, Bao L, Yang P, Chen D. Comparison of base excess, lactate and pH predicting 72-h mortality of multiple trauma. BMC Emerg Med. 2021;21:80.
19. Langer T, Brusatori S, Gattinoni L. Understanding base excess (BE): merits and pitfalls. Intensive Care Med. 2022;48:1080-3. Epub 2022 May 31.
20. Abbott TEF, Torrance HDT, Cron N, Vaid N, Emmanuel J. A single-centre cohort study of National Early Warning score (NEWS) and near patient testing in acute medical admissions. Eur J Intern Med. 2016;35:78-82. Epub 2016 Jun 23.
21. Caputo N, Fraser R, Paliga A, Kanter M, Hosford K, Madlinger R. Triage vital signs do not correlate with serum lactate or base deficit, and are less predictive of operative intervention in penetrating trauma patients: a prospective cohort study. Emerg Med J. 2013;30:546-50. Epub 2012 Jul 16.
22. Gupta S, Alam A. Shock index is better than conventional vital signs for assessing higher level of care and mortality in severe sepsis or shock. Am J Emerg Med. 2021;46:545-9. Epub 2020 Nov 10.
23. Heldeweg MLA, Langer T, Duška F. Guiding resuscitation in shock: base excess or lactate? Crit Care. 2024;28:249.
24. Hamed R, Mekki I, Aouni H, Hedhli H, Zoubli A, Maaref A, et al. Base excess usefulness for prediction of immediate mortality in severe trauma patients admitted to the emergency department. Tunis Med. 2019;97:1357-61.
25. Singhal R, Coghill JE, Guy A, Bradbury AW, Adam DJ, Scriven JM. Serum lactate and base deficit as predictors of mortality after ruptured abdominal aortic aneurysm repair. Eur J Vasc Endovasc Surg. 2005;30:263-6.
26. Caputo ND, Kanter M, Fraser R, Simon R. Comparing biomarkers of traumatic shock: the utility of anion gap, base excess, and serum lactate in the ED. Am J Emerg Med. 2015;33:1134-9. Epub 2015 May 7.
27. Callaway DW, Shapiro NI, Donnino MW, Baker C, Rosen CL. Serum lactate and base deficit as predictors of mortality in normotensive elderly blunt trauma patients. J Trauma. 2009;66:1040-4.
28. Gale SC, Kocik JF, Creath R, Crystal JS, Dombrovskiy VY. A comparison of initial lactate and initial base deficit as predictors of mortality after severe blunt trauma. J Surg Res. 2016;205:446-55. Epub 2016 Jul 5.
29. Zante B, Reichenspurner H, Kubik M, Kluge S, Schefold JC, Pfortmueller CA. Base excess is superior to lactate-levels in prediction of ICU mortality after cardiac surgery. PLoS One. 2018;13:e0205309.
30. Lu J, Dong Z, Ye L, Gao Y, Zheng Z. Predictive value of SOFA, PCT, Lactate, qSOFA and their combinations for mortality in patients with sepsis: a systematic review and meta-analysis. PLoS One. 2025;20:e0332525.
31. Davis JW, Dirks RC, Kaups KL, Tran P. Base deficit is superior to lactate in trauma. Am J Surg. 2018;215:682-5. Epub 2018 Jan 31.
32. Reese FB, Hubert FC, Cosentino MB, Oliveira MC, Réa Neto Á, Bernardelli RS, et al. Lactate and base excess (BE) as markers of hypoperfusion and mortality in traumatic hemorrhagic shock in patients undergoing damage control: a historical cohort. Rev Col Bras Cir. 2024;51:e20243699.
33. Candefjord S, Asker L, Caragounis EC. Mortality of trauma patients treated at trauma centers compared to non-trauma centers in Sweden: a retrospective study. Eur J Trauma Emerg Surg. 2022;48:525-36. Epub 2020 Jul 27.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.27003
Eurasian J Emerg Med. 2026;25: 186-93

Comparative Evaluation of Inflammatory, Nutritional and Composite Indices in Acute Pancreatitis

Selma Özlem Çelikdelen[1], Muhammet Cemal Kızılarslanoğlu[2]

[1]University of Health Sciences Türkiye, Konya City Hospital, Clinic of Internal Medicine, Konya, Türkiye
[2]University of Health Sciences Türkiye, Konya City Hospital, Clinic of Internal Medicine, Division of Geriatrics, Konya, Türkiye

Abstract

Aim: Early identification of patients with acute pancreatitis (AP) who are at risk of complicated disease and the need for intensive care unit (ICU) admission remains a major clinical challenge during emergency department evaluation and during inpatient follow-up. Easily accessible laboratory-based inflammatory and nutritional indices may provide practical tools for early risk stratification and clinical decision-making.

Materials and Methods: In this retrospective cohort study, 348 patients hospitalized with AP were evaluated. Inflammatory, nutritional, and composite indices derived from routine laboratory parameters obtained at hospital admission were calculated. Their associations with clinical outcomes were analyzed.

Results: Complicated pancreatitis occurred in 40.2% of patients and 10.3% of patients required ICU admission. Several inflammation-based indices, including the neutrophil-to-lymphocyte ratio, systemic immune-inflammation index, systemic inflammation response index and C-reactive protein-to-albumin ratio (CAR), were significantly associated with adverse outcomes. In multivariable analyses, CAR and the C-reactive protein-albumin-lymphocyte index (CALLY) remained independently associated with both complicated pancreatitis and the need for ICU admission. In ROC analysis, CALLY demonstrated the highest discriminative performance for predicting ICU requirement (area under the curve =0.808) with an optimal cut-off of≤0.05. Composite prognostic scores, including the Naples Prognostic score and Glasgow Prognostic score, were associated with outcomes in unadjusted analyses, but did not retain independent prognostic value.

Conclusion: Inflammation-based indices derived from routine laboratory tests particularly CAR and CALLY are associated with complicated pancreatitis and the need for ICU admission in patients with AP. Given their simplicity, low cost and availability at hospital admission, these indices may support early risk stratification and clinical management. Prospective multicenter studies are warranted to validate these findings.

Keywords: Acute pancreatitis, inflammatory indices, nutritional indices, C-reactive protein-to-albumin ratio

Introduction

Acute pancreatitis (AP) is a common gastrointestinal emergency with a highly variable clinical course, ranging from a mild, self-limiting disease to a severe condition associated with systemic inflammation, organ failure, and substantial morbidity (1). In the emergency department, AP represents a frequent cause of acute abdominal pain and often requires early decisions regarding hospitalization, the level of monitoring, and the need for intensive care. Despite advances in supportive care, early identification of patients at risk for a complicated disease course and for intensive care unit (ICU) admission remains a major clinical challenge. Timely risk stratification is crucial to optimize monitoring intensity and allocate critical care resources. It also plays a key role in guiding early therapeutic decisions, particularly during the initial assessment at hospital admission.

Several clinical scoring systems and imaging-based tools have been proposed to predict disease severity in AP; however, many are complex, time-consuming, or require repeated assessments, which may limit their practicality in time-sensitive emergency settings (2). In routine clinical practice, interest is growing in easily accessible laboratory-based biomarkers. These markers reflect the underlying inflammatory and nutritional status of the patient

Corresponding Author: Selma Özlem Çelikdelen MD, University of Health Sciences Türkiye, Konya City Hospital, Clinic of Internal Medicine, Konya, Türkiye
E-mail: drozlemkoc@hotmail.com **ORCID ID:** orcid.org/0009-0000-8546-0070

Cite this article as: Çelikdelen SÖ, Kızılarslanoğlu MC. Comparative evaluation of inflammatory, nutritional and composite indices in acute pancreatitis. Eurasian J Emerg Med. 2026;25: 186-93.

Received: 13.01.2026
Accepted: 11.02.2026
Published: 23.02.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

and can be readily obtained at presentation. Inflammatory indices derived from complete blood count parameters, such as the neutrophil-to-lymphocyte ratio (NLR) and related composite markers, have emerged as practical indicators of systemic inflammatory burden in various acute and chronic conditions (3). Beyond isolated inflammatory markers, integrated indices combining inflammatory and nutritional components such as the systemic immune-inflammation index (SII), systemic inflammation response index (SIRI), prognostic nutritional index (PNI), C-reactive protein-based ratios, and composite prognostic scores including the Naples Prognostic score (NPS) and Glasgow Prognostic score (GPS) have gained increasing attention. By integrating immune activation with nutritional reserve, these indices may better reflect the host response to acute inflammatory stress. This response represents a key determinant of outcomes in critically ill patients (4-8).

Although previous studies have explored selected inflammatory or nutritional markers in AP, comparative data evaluating a broad panel of inflammatory, nutritional and composite indices within the same cohort remain limited (9-12). Moreover, the relative performance of these indices in predicting clinically relevant outcomes particularly complicated pancreatitis and the need for ICU admission has not been fully clarified. Notably, composite prognostic scores such as the GPS and the NPS have not been extensively evaluated in patients with AP clarifying the prognostic utility of these readily available indices may improve early risk stratification in patients with AP.

Therefore, this study aimed to comprehensively evaluate the clinical relevance of multiple inflammatory and nutritional indices in patients hospitalized with AP. Specifically, we examined the associations of these indices with the development of complicated pancreatitis and the requirement for ICU admission, focusing on laboratory parameters available at hospital admission to identify readily accessible markers for early risk stratification in routine clinical practice.

Materials and Methods

Study Design and Patient Selection

This retrospective cohort study included adult patients (≥18 years) who were hospitalized for AP and followed in the internal medicine ward between March 2023 and September 2025. During this period, 372 patients were screened. All consecutive eligible patients presenting during the study period were evaluated for inclusion.

Patients were excluded if they were receiving routine hemodialysis (n=5), did not meet the Revised Atlanta Classification diagnostic criteria for AP (n=15), or had a concomitant malignancy on admission (n=4). After applying these exclusion criteria, a total of 348 patients were included in the final analysis.

Diagnosis and Classification of Acute Pancreatitis

The diagnosis of AP was confirmed according to the Revised Atlanta Classification and required the presence of at least two of the following three criteria: (1) characteristic abdominal pain consistent with AP; (2) serum amylase and/or lipase levels at least three times the upper limit of normal; and (3) imaging findings compatible with AP (13).

Disease severity was classified as mild, moderately severe, or severe based on the Revised Atlanta Classification. The presence of complicated pancreatitis, including local and/or systemic complications, was also determined in accordance with the same classification system (13).

Data Collection

Demographic data (including age and sex), body mass index, etiology of pancreatitis, and comorbidity burden were obtained from electronic medical records. Comorbid conditions were quantified using the Charlson Comorbidity index (CCI), a validated scoring system that assigns weighted scores to predefined chronic diseases, as described by Quan et al. (14). Laboratory parameters were obtained from blood samples collected at hospital admission, prior to initiation of disease-specific treatment, and reflect routine laboratory evaluation performed during the initial presentation to the emergency department.

Inflammatory and Nutritional Indices

Inflammatory, nutritional, and composite indices were calculated using laboratory values obtained at admission. The evaluated indices included NLR, platelet-to-lymphocyte ratio (PLR), lymphocyte-to-monocyte ratio (LMR), SII, SIRI, PNI, C-reactive protein-to-albumin ratio (CAR), C-reactive protein-albumin-lymphocyte index (CALLY), NPS, and GPS. All indices were calculated according to previously published definitions (3-8).

The NPS was calculated based on serum albumin, total cholesterol, NLR, and LMR, with patients categorized according to established scoring criteria. The GPS was determined using C-reactive protein and serum albumin levels, as previously described.

Clinical Outcomes

The predefined clinical outcomes of interest were the development of complicated pancreatitis and the requirement for ICU admission during hospitalization. Both outcomes were recorded based on the clinical course during follow-up.

Ethical approval for this study was obtained from the Ethics Committee of Karatay University Faculty of Medicine with the

(approval number: 2025/043, date: 30.10.2025). The study was conducted in accordance with the principles of the World Medical Association Declaration of Helsinki (revised in 2013). Our study was conducted retrospectively through review of the electronic medical records, laboratory findings, and imaging reports of hospitalized patients. No direct contact with patients was made, and no interventions were performed. Therefore, individual informed consent was not required.

Statistical Analysis

Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS) software, version 26.0 (IBM Corp., Armonk, NY, USA). Categorical variables were expressed as counts and percentages. Continuous variables were presented as mean ± standard deviation for normally distributed data and as median (minimum-maximum) for non-normally distributed data. Comparisons between groups were conducted using the chi-square test or Fisher's exact test for categorical variables, and the Student's t-test or Mann-Whitney U test for continuous variables, as appropriate.

The associations of inflammatory and nutritional indices with the development of complicated pancreatitis and with ICU requirement were evaluated using univariate and multivariable logistic regression analyses. Multivariable models were adjusted for clinically relevant covariates, including age, sex, etiology of pancreatitis, and selected biochemical parameters. Results were reported as odds ratios with 95% confidence intervals.

ROC analyses were performed to assess the discriminative performance of inflammatory and nutritional indices in predicting complicated pancreatitis and ICU requirement; AUCs were calculated. A p value <0.05 was considered statistically significant.

Results

Baseline Characteristics: A total of 348 patients with AP were included; 196 (56.3%) were female. The median age was 58 years (range 19-97). The median CCI score was 0 (0-4) and the median length of hospital stay was 6 days (3-40). Biliary etiology accounted for 181 cases (52.0%) while 167 cases (48.0%) were of non-biliary origin. Disease severity was classified as mild in 204 (58.6%) patients, moderate in 120 (34.5%) patients, and severe in 24 (6.9%) patients. ICU admission was required for 36 patients (10.3%). Complicated pancreatitis occurred in 140 patients (40.2%), and overall mortality was 0.9% (n=3). Baseline demographic, clinical, and laboratory characteristics, including inflammatory and nutritional indices, are summarized in Table 1.

Table 1. General characteristics of the patients

Parameters	n (%) or mean ± SD or median (minimum-maximum)
General features	
Total patient numbers	348 (100.0)
Gender, female	196 (56.3)
Age, years	58 (19-97)
BMI, kg/m^2	25.9 (17.5-44.6)
Older patients, age ≥65 years	133 (38.2)
CCI	0 (0-4)
Length of hospital stay, days	6 (3-40)
Characteristics of pancreatitis	
Etiology of pancreatitis	
Nonbiliary origin	167 (48.0)
Biliary origin	181 (52.0)
Severity of pancreatitis	
Mild	204 (58.6)
Moderate	120 (34.5)
Severe	24 (6.9)
ICU requirement	36 (10.3)
Complicated pancreatitis	140 (40.2)
Mortality	3 (0.9)
Indices	
NLR	5.66 (0.80-89.12)
SII	1448.9 (139.6-31636.0)
SIRI	3.92 (0.32-83.64)
PNI	47.89±7.71
PLR	166.5 (37.2-2100.0)
CAR	0.36 (0.01-9.74)
CALLY	0.38 (0.00-40.95)
LMR	2.19 (0.21-16.67)
NPS	3 (0-4)
GPS	1 (0-2)

Data are presented as n (%), mean ± standard deviation, or median (minimum-maximum), as appropriate
BMI: Body mass index, CCI: Charlson Comorbidity index, ICU: Intensive care unit, NLR: Neutrophil-to-lymphocyte ratio, SII: Systemic immune-inflammation index, SIRI: Systemic inflammation response index, PLR: Platelet-to-lymphocyte ratio, CAR: C-reactive protein-to-albumin ratio, CALLY: C-reactive protein-albumin-lymphocyte index, LMR: Lymphocyte-to-monocyte ratio, PNI: Prognostic nutritional index, NPS: Naples prognostic score, GPS: Glasgow Prognostic score, SD: Standard deviation

Comparison of Inflammatory and Nutritional Indices According to Clinical Outcomes: Patients with complicated pancreatitis exhibited a significantly higher inflammatory burden compared with those with uncomplicated pancreatitis. Median values of NLR, SII, SIRI, PLR, and CAR were markedly elevated, whereas nutritional indices, including PNI and CALLY, were significantly lower in the complicated pancreatitis group (all p<0.001). Similarly, NPS and GPS scores were modestly but significantly higher among patients with complications.

A consistent pattern was observed in analyses stratified by ICU requirement. Patients requiring ICU care had substantially higher levels of NLR, SII, SIRI, PLR, and CAR, and lower values of PNI, CALLY, and LMR compared with non-ICU patients (all p≤0.005). Additionally, both the NPS and GPS scores were significantly higher in patients admitted to the ICU. These findings are summarized in Table 2.

Multivariable Logistic Regression Analysis

In multivariable models adjusted for age, sex, etiology of pancreatitis, and relevant biochemical parameters, several inflammatory indices remained independently associated with adverse clinical outcomes. In complicated pancreatitis, higher NLR, SII, SIRI, and CAR levels were independently associated with an increased risk of complications, whereas higher CALLY values were inversely associated with the development of complications. Among these parameters, CAR demonstrated the strongest independent association with complicated pancreatitis.

Similarly, in analyses evaluating ICU requirement, elevated levels of NLR, SII, SIRI, PLR, and CAR remained independently associated with ICU requirement after adjustment. In contrast, PNI and LMR, as well as the composite scores NPS and GPS, did not retain independent significance in adjusted models. These findings are summarized in Table 3.

ROC Curve Analyses

ROC curve analyses were performed to evaluate the discriminatory performance of inflammatory and nutritional indices in predicting complicated pancreatitis and the need for ICU admission. In complicated pancreatitis, inflammatory indices showed modest discriminatory performance, with the highest AUC values observed for CALLY, NLR and SIRI, while composite prognostic scores demonstrated limited discrimination. Based on ROC-derived optimal cut-off values, CALLY ≤0.20 and CAR >0.60 were associated with an increased risk of complicated pancreatitis and showed moderate sensitivity and specificity (Table 4).

Table 2. Comparison of laboratory indices between groups based on pancreatitis complications and ICU requirements

Indices	**Uncomplicated pancreatitis**	**Complicated pancreatitis**	**p value**
NLR	4.29 (0.8-45.88)	8.66 (1.02-89.12)	<0.001
SII	1071.94 (139.57-11457.95)	2053.96 (156.63-31635.96)	<0.001
SIRI	2.96 (0.32-65.46)	5.31 (0.48-83.64)	<0.001
PNI	49.24±7.11	45.9±8.16	<0.001
PLR	148.65 (43.62-1015.38)	204.45 (37.21-2100)	<0.001
CAR	0.26 (0.01-5.06)	0.64 (0.01-9.74)	<0.001
CALLY	0.56 (0.01-40.95)	0.18 (0-10.32)	<0.001
LMR	2.43 (0.28-16.67)	1.9 (0.21-12.5)	<0.001
NPS*	3 (0-4)	3 (0-4)	0.012
GPS*	1 (0-2)	1 (0-2)	<0.001
Indices	**No ICU requirement**	**ICU requirement**	**p value**
NLR	5.05 (0.8-68.56)	13.79 (2.08-89.12)	<0.001
SII	1336.32 (139.57-14356.36)	3384.91 (592.7-31635.96)	<0.001
SIRI	3.67 (0.32-83.64)	9.56 (0.86-68)	<0.001
PNI	48.45±7.06	42.94±10.91	0.005
PLR	159.46 (40.37-1021.21)	235.81 (37.21-2100)	<0.001
CAR	0.31 (0.01-8.29)	1.88 (0.12-9.74)	<0.001
CALLY	0.47 (0.01-40.95)	0.04 (0-5.87)	<0.001
LMR	2.25 (0.22-16.67)	1.39 (0.21-12.5)	0.002
NPS	3 (0-4) and (2.69±0.99)	3 (2-4) and (3.22±0.75)	0.005
GPS	1 (0-2) and (0.67±0.61)	1 (0-2) and (1.11±0.71)	<0.001

Data are presented as median (minimum-maximum) or mean ± standard deviation, as appropriate.
*For additional information, mean ± standard deviation values were as follows:
NPS: 2.64±1.02 in uncomplicated pancreatitis and 2.92±0.91 in complicated pancreatitis;
GPS: 0.62±0.62 in uncomplicated pancreatitis and 0.86±0.62 in complicated pancreatitis
NLR: Neutrophil-to-lymphocyte ratio, SII: Systemic immune-inflammation index, SIRI: Systemic inflammation response index, PLR: Platelet-to-lymphocyte ratio, CAR: C-reactive protein-to-albumin ratio, CALLY: C-reactive protein-albumin-lymphocyte index, LMR: Lymphocyte-to-monocyte ratio, PNI: Prognostic nutritional index, NPS: Naples prognostic score, GPS: Glasgow prognostic score, ICU: Intensive care unit

Table 3. Multiple logistic regression models showing associations of the laboratory indices with complicated pancreatitis and ICU requirement

Parameters	Unadjusted		Adjusted	
	OR	95% CI	OR	95% CI
Complicated pancreatitis				
NLR	1.065***	1.036-1.095	1.037*	1.005-1.069
SII	1.029***	1.017-1.041	1.021**	1.007-1.036
SIRI	1.072***	1.036-1.110	1.042*	1.002-1.084
PNI	0.942***	0.913-0.971	0.995	0.951-1.040
PLR	1.003***	1.001-1.004	1.001	1.000-1.003
CAR	1.764***	1.408-2.211	1.841***	1.393-2.443
CALLY	0.771**	0.658-0.904	0.845*	0.731-0.976
LMR	0.839*	0.730-0.964	0.900	0.777-1.041
NPS	1.364*	1.069-1.741	1.177	0.863-1.604
GPS	1.843**	1.295-2.624	1.287	0.839-1.975
ICU requirement				
NLR	1.075***	1.046-1.104	1.063**	1.026-1.101
SII	1.028***	1.017-1.039	1.027***	1.012-1.042
SIRI	1.069***	1.037-1.102	1.057**	1.014-1.101
PNI	0.907***	0.864-0.952	0.979	0.909-1.055
PLR	1.003***	1.002-1.005	1.003**	1.001-1.005
CAR	1.836***	1.496-2.253	1.961***	1.481-2.596
CALLY	0.387*	0.187-0.801	0.675	0.365-1.249
LMR	0.853	0.670-1.086	0.965	0.764-1.221
NPS	1.891**	1.204-2.968	1.308	0.730-2.342
GPS	3.047***	1.731-5.365	1.782	0.873-3.637

*Indicates p values <0.05, **indicates p value <0.01, and ***indicates p value <0.001
All indices were adjusted for age, gender, etiology of pancreatitis, and levels of triglycerides, calcium, bilirubin, lactate dehydrogenase, albumin, and creatinine, except for the CAR, PNI, CALLY, and GPS scores, which were adjusted using the same parameters except for albumin.
The SII levels were included in the regression models as SII/100
NLR: Neutrophil-to-lymphocyte ratio, SII: Systemic immune-inflammation index, SIRI: Systemic inflammation response index, PLR: Platelet-to-lymphocyte ratio, CAR: C-reactive protein-to-albumin ratio, CALLY: C-reactive protein-albumin-lymphocyte index, LMR: Lymphocyte-to-monocyte ratio, PNI: Prognostic nutritional index, NPS: Naples prognostic score, GPS: Glasgow Prognostic score, ICU: Intensive care unit, OR: Odds ratio, CI: Confidence interval

In analyses assessing the need for ICU care, several indices demonstrated greater discriminatory performance. CALLY demonstrated the highest accuracy (AUC =0.808), followed by SII (AUC =0.784), CAR (AUC =0.781), NLR (AUC =0.771), and SIRI (AUC =0.761). The optimal CALLY cut-off for ICU requirement was ≤0.05, with high specificity (94.8%) and negative predictive value (94.8%) (Table 4).

Overall, inflammatory indices showed moderate-to-good discrimination for ICU requirement, whereas composite scores (NPS and GPS) showed modest discrimination. The ROC analysis results are summarized in Table 4 and the corresponding ROC curves are shown in Figure 1.

Discussion

In this retrospective cohort study, we comprehensively evaluated a broad panel of inflammatory, nutritional, and composite indices obtained at hospital admission from patients with AP. Our findings indicate that several inflammation-based indices were associated with the development of complicated pancreatitis and the need for ICU admission; however, only a subset of these indices remained independently associated with clinical outcomes in multivariable analyses.

Consistent with previous literature, indices of systemic inflammatory burden, including NLR, SII, SIRI, and CAR, were elevated in patients who developed complicated pancreatitis or required ICU care. In multivariable analyses, several of these

Table 4. ROC curve analysis of laboratory indices for complicated pancreatitis and ICU requirement

Parameters	AUC	Cut-off	p value	Sensitivity %	Specificity %	PPV %	NPV %
For complicated pancreatitis							
NLR	0.686	>5.1	<0.001	72.9	58.9	54.5	76.2
SII	0.676	>16.8	<0.001	61.4	71.0	58.9	73.1
SIRI	0.681	>3.5	<0.001	72.1	57.9	53.7	75.5
PLR	0.615	>178.2	0.001	60.7	66.2	54.8	71.4
CAR	0.662	>0.6	<0.001	52.1	72.1	55.7	69.1
NPS	0.585	>2.0	0.013	73.2	42.3	46.2	70.0
GPS	0.583	>0.0	0.015	72.5	45.6	47.2	71.2
PNI	0.633	≤47.0	<0.001	58.6	65.2	53.2	69.9
CALLY	0.690	≤0.20	<0.001	54.3	76.8	61.3	71.3
LMR	0.617	≤1.7	<0.001	47.9	75.4	56.8	68.1
For ICU requirement							
NLR	0.771	>8.4	<0.001	77.8	69.7	23.0	96.4
SII	0.784	>15.8	<0.001	88.9	59.0	20.1	97.9
SIRI	0.761	>8.4	<0.001	58.3	82.3	27.6	94.4
PLR	0.682	>155.9	0.001	86.1	48.7	16.3	96.8
CAR	0.781	>1.0	<0.001	69.4	78.8	27.5	95.7
NPS	0.640	>2.0	0.010	81.3	38.2	13.4	94.5
GPS	0.647	>1.0	0.007	30.7	92.8	33.3	89.5
PNI	0.711	≤43.9	<0.001	58.3	75.8	21.1	93.6
CALLY	0.808	≤0.05	<0.001	55.6	94.8	55.6	94.8
LMR	0.656	≤1.6	0.002	61.1	72.3	20.4	94.1

Optimal cut-off values were determined using ROC curve analysis based on the Youden index. Sensitivity, specificity, PPV, and NPV were calculated accordingly
NLR: Neutrophil-to-lymphocyte ratio, SII: Systemic immune-inflammation index, SIRI: Systemic inflammation response index, PLR: Platelet-to-lymphocyte ratio, CAR: C-reactive protein-to-albumin ratio, CALLY: C-reactive protein-albumin-lymphocyte index, LMR: Lymphocyte-to-monocyte ratio, PNI: Prognostic nutritional index, NPS: Naples prognostic score, GPS: Glasgow Prognostic score, ICU: Intensive care unit, PPV: Positive predictive value, NPV: Negative predictive value, AUC: Area under the curve

indices remained independently associated with adverse clinical outcomes. Wu et al. (9) demonstrated a significant association between SIRI and disease severity in AP, underscoring the role of systemic inflammation in disease progression. Similarly, Khanal et al. (11) reported that the CAR is a useful predictor of disease severity in AP. In addition, inflammation-based models incorporating multiple inflammatory markers have been shown to aid in the identification of severe AP (15). Similarly, our study demonstrated that these inflammation-based indices were associated with clinically relevant indicators of disease severity in AP. These included the development of complicated pancreatitis and the need for ICU admission.

Among all evaluated parameters, the CALLY index emerged as a particularly informative marker. Higher CALLY values were inversely associated with the development of complicated pancreatitis and demonstrated the highest discriminative performance for predicting ICU requirement in our cohort. This finding is consistent with the study by Xu et al. (12), who identified CALLY as a strong predictor of severe AP in a large single-center retrospective cohort. By combining inflammatory and nutritional components, CALLY may provide a more comprehensive reflection of the host response to acute inflammatory stress.

The PNI and the LMR were significantly lower in patients with complicated disease and in those requiring ICU care; however, neither parameter retained independent significance in adjusted models. These findings are consistent with Genc et al. (10), who reported that PNI was not significantly associated with overall complication status, but was lower in patients who developed pancreatic necrosis. This suggests that while nutritional status may be related to specific disease manifestations, systemic inflammatory burden appears to play a more prominent role in determining early clinical outcomes in AP.

Composite prognostic scores, including the NPS and GPS, were modestly associated with adverse outcomes in unadjusted analyses, but did not demonstrate independent predictive value in multivariable models. Data regarding the prognostic role of these composite scores in AP remain limited. To our

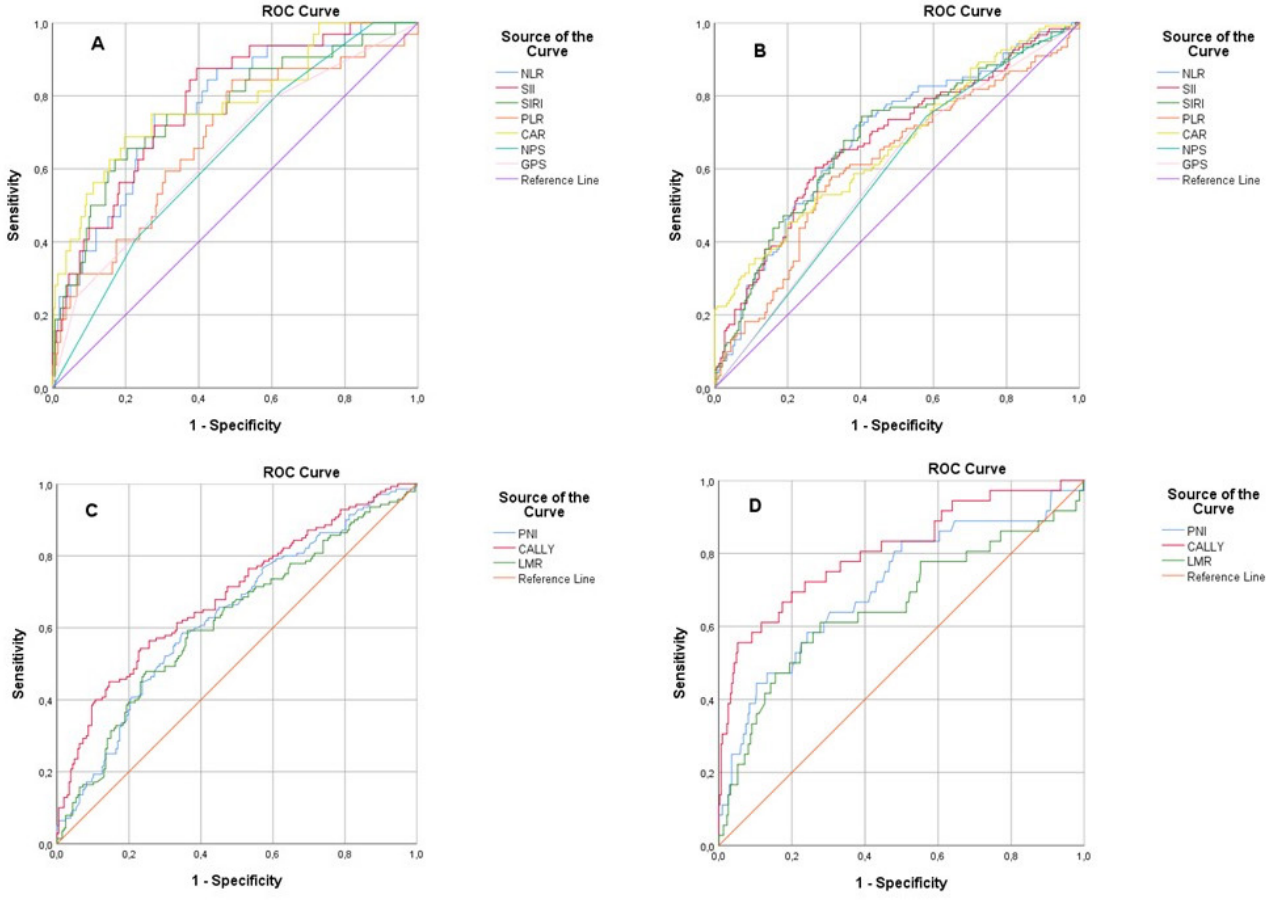


Figure 1. ROC curve analyses of inflammatory and nutritional indices for predicting complicated pancreatitis and intensive care unit (ICU) requirement. Panels A and D illustrate ROC curves for ICU requirement, while panels B and C illustrate ROC curves for complicated pancreatitis. AUC values are provided for each index

NLR: Neutrophil-to-lymphocyte ratio, SII: Systemic immune-inflammation index, SIRI: Systemic inflammation response index, PLR: Platelet-to-lymphocyte ratio, CAR: C-reactive protein-to-albumin ratio, CALLY: C-reactive protein-albumin-lymphocyte index, LMR: Lymphocyte-to-monocyte ratio, PNI: Prognostic nutritional index, NPS: Naples prognostic score, GPS: Glasgow Prognostic score, ICU: Intensive Care unit, PPV: Positive predictive value, NPV: Negative predictive value, AUC: Area under the curve

knowledge, this is one of the first studies to comprehensively evaluate NPS and GPS in this clinical setting. Our findings suggest that while these scores may reflect overall health status and inflammatory burden, their standalone prognostic utility for early risk stratification in AP appears limited when established clinical variables and systemic inflammatory markers are taken into account.

Study Limitations

Several limitations of this study should be acknowledged. First, the retrospective, and single-center design may limit the generalizability of the findings. Second, laboratory indices were assessed only at hospital admission; dynamic changes during hospitalization were not evaluated. Third, external validation in independent cohorts was not performed. Despite these limitations, the strengths of the study include a relatively large sample size and a comprehensive evaluation of multiple indices within the same cohort. In addition, the focus on clinically meaningful outcomes using laboratory parameters that are readily available at initial presentation enhances the relevance of our findings to routine emergency and acute-care practice.

From an emergency medicine perspective, the ability to identify high-risk patients using routinely available laboratory parameters at the time of hospital admission is of particular importance. Early risk stratification may inform triage decisions, guide monitoring intensity, and facilitate timely ICU referral. In this regard, indices such as CALLY, CAR and SIRI which demonstrated consistent associations with clinically relevant outcomes may serve as practical tools for emergency physicians during the initial assessment of patients AP.

Conclusion

In conclusion, this study demonstrates that inflammation-based indices derived from routine laboratory tests particularly the CAR and CALLY indices are associated with progression to complicated pancreatitis and with the need for intensive care among patients with AP. Owing to their simplicity, low cost, and availability at hospital admission, these indices may serve as practical tools for early risk stratification in the emergency setting, facilitating timely triage decisions between ward and ICU. Furthermore, by aiding the early identification of high-risk patients, these markers may support clinicians in optimizing monitoring strategies and guiding appropriate clinical management. Prospective, multicenter studies are warranted to validate these findings and further clarify their role in routine clinical practice.

Ethics

Ethics Committee Approval: Ethical approval for this study was obtained from the Ethics Committee of Karatay University Faculty of Medicine with the (approval number: 2025/043, date: 30.10.2025). The study was conducted in accordance with the principles of the World Medical Association Declaration of Helsinki (revised in 2013).

Informed Consent: Our study was a retrospective review of the electronic medical records, laboratory findings, and imaging reports of hospitalized patients. No direct contact with the patients was made, and no interventions were performed. Therefore, individual informed consent was not required.

Footnotes

Authorship Contributions

Concept: S.Ö.Ç., M.C.K., Design: S.Ö.Ç., M.C.K., Data Collection or Processing: S.Ö.Ç., Analysis or Interpretation: M.C.K., Literature Search: S.Ö.Ç., Writing: S.Ö.Ç.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Wang CF, Tariq A, Chandra S. Acute pancreatitis. 2025 Aug 2. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025.
2. Zhu J, Wu L, Wang Y, Fang M, Liu Q, Zhang X. Predictive value of the ranson and BISAP scoring systems for the severity and prognosis of acute pancreatitis: a systematic review and meta-analysis. PLoS One. 2024;19:e0302046.
3. Pintea-Trifu ML, Balici SŞ, Vică ML, Leucuța DC, Coman HG, Nemeș B, et al. Neutrophil to lymphocyte ratio, platelet to lymphocyte ratio, lymphocyte to monocyte ratio and systemic inflammatory index in sexually transmitted diseases. Med Pharm Rep. 2024;97:162-8. Epub 2024 Apr 25.
4. Hu B, Yang XR, Xu Y, Sun YF, Sun C, Guo W, et al. Systemic immune-inflammation index predicts prognosis of patients after curative resection for hepatocellular carcinoma. Clin Cancer Res. 2014;20:6212-22. Epub 2014 Sep 30.
5. Qi Q, Zhuang L, Shen Y, Geng Y, Yu S, Chen H, et al. A novel systemic inflammation response index (SIRI) for predicting the survival of patients with pancreatic cancer after chemotherapy. Cancer. 2016;122:2158-67. Epub 2016 May 6.
6. Bekki T, Shimomura M, Yano T, Watanabe A, Ishikawa S, Imaoka K, et al. C-reactive protein-albumin-lymphocyte index is a useful indicator for recurrence and survival following curative resection of stage i-iii colorectal cancer. J Anus Rectum Colon. 2025;9:192-201.
7. Kang Y, Tan Y, Tie Y, Zhang Y, Cui C, Zhao Y. Prognostic value of the naples prognostic score in adult chronic obstructive pulmonary disease: NHANES 2005-2018. Front Nutr. 2024;11:1502266.
8. Nozoe T, Matono R, Ijichi H, Ohga T, Ezaki T. Glasgow prognostic score (GPS) can be a useful indicator to determine prognosis of patients with colorectal carcinoma. Int Surg. 2014;99:512-7.
9. Wu W, Zhang Y, Zhang Y, Qu X, Zhang Z, Zhang R. Association between the systemic inflammation response index and severity of acute pancreatitis: a retrospective cohort study. J Inflamm Res. 2025;18:4471-80.
10. Genc M, Duran EN, Ekinci I, Bayram M, Uzun H, Tabak O. Association of nutritional status indices with clinical outcomes in acute pancreatitis. Front Nutr. 2025;12:1662958.
11. Khanal N, Baral A, Shrestha S, Lakhey PJ, Bhandari RS. Predictive value of c-reactive protein/albumin ratio in acute pancreatitis. Cureus. 2025;17:e93436.
12. Xu X, Zou Y, Ke H, Kuang M, Xiong S, Ding L, et al. Association of the c-reactive protein-to-albumin-to-lymphocyte index with severe acute pancreatitis: a single-center retrospective study of 5,016 patients. Pancreatology. 2026;26:40-8. Epub 2025 Dec 11.
13. Banks PA, Bollen TL, Dervenis C, Gooszen HG, Johnson CD, Sarr MG, et al. Classification of acute pancreatitis--2012: revision of the Atlanta classification and definitions by international consensus. Gut. 2013;62:102-11. Epub 2012 Oct 25.
14. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and validating the charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol. 2011;173:676-82. Epub 2011 Feb 17.
15. Li X, Zhang Y, Wang W, Meng Y, Chen H, Chu G, et al. An inflammation-based model for identifying severe acute pancreatitis: a single-center retrospective study. BMC Gastroenterol. 2024;24:63.7

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.22308
Eurasian J Emerg Med. 2026;25: 194-8

Trauma Team Activation at an Emergency Department

Sze Jia Lam[1], Hersh Attal[1], Weei Jang Daniel Tan[1], Kuhan Venugopal[1], Shiun-hwa Chantal Lim[2], Chee Meng Lee[3], Jen Heng Pek[1], Yee Lyn Corinne Lau[1]

[1]Sengkang General Hospital, Clinic of Emergency Medicine, Singapore
[2]Sengkang General Hospital, Division of Hyperacute Care, Singapore
[3]Sengkang General Hospital, Clinic of Surgery, Singapore

Abstract

Aim: Multidisciplinary trauma teams are routinely activated for patients with serious injuries in the emergency department (ED). We aim to describe the characteristics of patients requiring trauma team activation (TTA) at a level two trauma centre.

Materials and Methods: A retrospective review of a single centre's trauma registry data was performed. Information on demographics, circumstances of trauma, details of injury, and clinical progress was collected and analysed.

Results: Among 77,132 trauma cases, 496 (0.6%) required TTA. On average, one TTA occurred every three days. The median age was 38 years [interquartile range (IQR), 28-54 years], and 367 (74.0%) were male. The primary reason for TTA was the mechanism of injury (n=418, 84.3%). Blunt injuries occurred in 453 (91.3%) patients and vehicular accidents were the most common mechanism (n=342, 69.0%). The median injury severity score was 1 (IQR: 0-9). Twenty patients (4.0%) required emergency surgery. The overall mortality was 60 (12.1%), with 49 patients (9.9%) dying in the ED. Four hundred and one patients (80.8%) were admitted to the hospital; the median length of stay was 2 days (IQR 1-5 days).

Conclusion: TTA was an uncommon event; when it occurred, patients generally had low injury severity and low utilisation of healthcare resources. Revision of the activation criteria and a tiered trauma-team or expedited-care approach may help achieve a balance between timely, coordinated care and the justified, optimal deployment of personnel.

Keywords: Emergency department, multi-disciplinary, trauma

Introduction

A trauma team is a multidisciplinary group of healthcare professionals that can be activated using predefined criteria to attend to patients with serious injuries in the emergency department (ED). The primary aims of the team are to resuscitate and stabilise the patient, to determine the extent and severity of the injuries, and to prioritise the management for timely definitive care (1). By working as a team with a leader and members who have specific roles and responsibilities, the trauma team can reduce the time to diagnosis and treatment, optimise resource utilisation, and improve patient outcomes (2).

The activation criteria should be tailored to the local setting of the trauma centre to avoid undertriage and overtriage. According to the American College of Surgeons' Committee on Trauma, acceptable rates of under-triage range from 1-5%, and those for over-triage range from 25-35% (3). Under-triage is defined as patients who were not trauma-activated but required high-level resources, whereas over-triage is defined as patients who were trauma-activated but did not require high-level resources. Inappropriate use of trauma team activation (TTA) can lead to overutilisation of valuable personnel when trauma team members are called to the ED, disrupting their duties to provide care in other areas of the hospital (4,5). Therefore, our study aimed to describe the characteristics of patients requiring TTA at the ED, so that the information gained from this work may be used to review the activation criteria and optimise manpower deployment without compromising the care of patients with serious injuries.

Corresponding Author: Jen Heng Pek MD, Sengkang General Hospital, Clinic of Emergency Medicine, Singapore
E-mail: jenheng_@hotmail.com **ORCID ID:** orcid.org/0000-0002-8356-7410
Cite this article as: Lam SJ, Attal H, Tan WJD, Venugopal K, Lim SHC, Lee CM et al. Trauma team activation at an emergency department. Eurasian J Emerg Med. 2026;25: 194-8.

Received: 05.12.2025
Accepted: 11.02.2026
Published: 23.02.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Materials and Methods

Study Setting

Our study was conducted at the ED of a level-two trauma centre in Singapore, with an annual census of about 122,000; trauma patients accounted for approximately 15% of visits. This level-two trauma centre is part of a regional trauma system comprising two other level-one trauma centres.

Pre-hospital emergency medical services (EMS) are provided by the Singapore Civil Defence Force, and patients are routinely conveyed from the incident site to the nearest ED. However, trauma patients meeting diversion criteria (Table 1) will be transferred only to level-one trauma centres, unless they are in extremis due to cardiorespiratory arrest. Our hospital has a single-tier trauma team that consists primarily of emergency doctors and nurses, general surgeons, and orthopaedic surgeons. The team is activated based on the criteria in Table 2 and will be led by the general surgeon. When required, additional personnel from relevant departments, such as neurosurgery, will be called upon to assist with patient care. Transfer to a level-one trauma centre in the regional trauma system will be undertaken if further trauma expertise available only at that centre is required. Otherwise, patients may be admitted to the hospital under the appropriate specialty for further management.

Study Design

Our study was a retrospective review of the trauma registry data that included all trauma patients presenting to the ED at a single level-two trauma centre from 1 August 2018 to 31 December 2022. We collected and analysed information, including demographics, types and mechanisms of injury, injuries sustained and their severities, investigations, treatments, and outcomes.

The evaluation of under- and over-triage according to TTA criteria was based on the utilisation of high-level resources by trauma patients. This definition of high-level resources included the following ED interventions: intubation, chest tube insertion, intraosseous insertion, central line insertion, blood administration, cardiopulmonary resuscitation, pericardiocentesis, open thoracotomy, and time to operation within 90 minutes of hospital arrival (6). The proportion of under-triage was determined as the number of trauma patients who utilised high-level resources but did not have TTA, divided by the total number of trauma patients in the study, while the proportion of over-triage was determined as the number of trauma patients who did not utilise high-level resources but had TTA, divided by the total number of trauma patients in the study.

This study was approved by Institutional Review Board of SingHealth, Singapore (CRIB reference: 2021/2417, date: 03.07.2021).

Statistical Analysis

Statistical analysis was performed using SPSS version 25 (SPSS, Chicago, IL). Continuous data were presented as median with interquartile range (IQR). Categorical data were presented as frequencies and percentages.

Results

A total of 77,132 trauma patients were seen in the ED during the study period. Of these, 496 patients (0.6%) were managed by the trauma team, corresponding to an average of one TTA every three days.

Characteristics of Trauma Patients with TTA at the ED

The median patient age was 38 years (IQR 28-54), and 367 (74.0%) of them were male. Seven (1.4%) patients were younger than 16 years, and 73 (14.7%) patients were 65 years or older. TTA was due to mechanism-of-injury, physiologic, and anatomic criteria in 418, 91, and 93 patients, respectively. Eighty five patients met more than one criterion for TTA. Blunt injuries occurred in 453 (91.3%) patients, and vehicular accidents were the most common mechanism of injury (n=342, 69.0%) (Table 3).

Outcomes of Trauma Patients with TTA at the ED

Among 414 (83.5%) patients admitted to the hospital, 333 were admitted to the general ward, 62 to the high-dependency unit, and 19 to the intensive care unit. Twenty (4.0%) patients underwent emergency surgery following direct operating theatre transfer from the ED. The overall median length of stay for admitted patients was 2 days (IQR 1-5). Sixty patients (12.1%) died: 49 arrived in cardiorespiratory arrest and died in the ED, and 11 died subsequently during the hospital stay. Excluding these patients who died because their cause of death and extent of injuries were not available in the coroner's report, the median injury severity score (ISS) of the remaining patients was 1 (IQR 0 to 9). Of these remaining patients, 202 had an ISS of 0; 196 had an ISS of 1 to 15, which is suggestive of minor trauma; and 38 had an ISS of 16 or more, which is suggestive of major trauma.

Twenty patients (4.0%) required transfer to another hospital: 14 were transferred to a level-one trauma centre, four to a hospital with a paediatric unit, and two to a hospital with a burn unit. Fourteen patients (2.8%) were discharged from the ED.

Performance of the Criteria for TTA

408 patients who were trauma-activated did not require high-level resources, whereas 63 patients who were trauma-

Table 1. Criteria for pre-hospital EMS diversion of a trauma patient to a level one trauma centre
Physiology Systolic blood pressure less than 90 mmHg Glasgow Coma scale score 13 or less
Anatomy Penetrating injury to head, neck, torso and extremities proximal to elbow or knee Chest wall instability or deformity Two or more proximal long bone fractures Crushed, degloved, mangled or pulseless extremity (includes near amputation) Amputation proximal to wrist or ankle Pelvic fracture Open or depressed skull fracture Spinal cord injury
Mechanism Fall from height more than 3 meters High risk vehicle crash-ejection from vehicle, death from same vehicle, vehicle intrusion more than 30 cm in occupant site Motorcycle crash more than 50 km/h Vehicle crash with pedestrian/cyclist-more than 30 km/h or thrown or run over Trapped body or limb between a structure, machinery or vehicle more than 20 minutes High voltage injury Explosion

Table 2. Criteria for trauma team activation at the emergency department
Physiology Airway obstruction Respiratory rate less than 10 or more than 30 per minute Systolic blood pressure less than 90 mmHg or pulse rate more than 120 per minute Glasgow Coma scale score less than 13 Pregnant
Anatomy Penetrating injury to head, neck, torso Flail chest Two or more body region injury Two or more proximal long bone fractures Proximal limb amputation Pelvic fracture Open or depressed skull fracture Spinal cord injury
Mechanism Fall from height more than two meters Road traffic accident more than 50 km/h Ejection from vehicle or flung by vehicle Death from same vehicle Prolonged entrapment more than 20 minutes Crush injury High voltage injury Explosion

activated required high-level resources. These corresponded to under-triage and over-triage proportions of 0.08% and 82.3%, respectively. Among patients who were under-triaged, the high-level resources utilised were endotracheal intubation, tube thoracostomy, and blood administration.

Discussion

Unlike a level one trauma centre, TTA was uncommon in the ED of a level-two trauma centre. Here, TTA was commonly based on mechanism-of-injury criteria and, correspondingly, the ISS of patients, the utilisation of healthcare resources, and the proportion of patients who died during the inpatient hospital stay were low. Therefore, to reduce the imbalance between under- and over-triage proportions in the criteria for TTA, revision of the activation criteria and implementation of a tiered trauma team or expedited-care approach may help achieve a balance between timely, coordinated care and the justified, optimal deployment of manpower in a level-two trauma centre.

Table 3. Circumstances of injuries	
Variable	**n (%)**
Type of injury	
Blunt	453 (91.3)
Penetrating	43 (8.7)
Mechanism of Injury	
Vehicular accident	342 (69.0)
Fall	101 (20.4)
Tool/object	25 (5.0)
Interpersonal violence	11 (2.2)
Sports	4 (0.8)
Others	4 (0.8)
Unknown	9 (1.8)

The mechanism of injury as a criterion for TTA has been associated with high rates of over-triage (7,8). The circumstances of injury raise clinical concern for an underlying injury, which mandates prompt and thorough evaluation. However, there may be no injuries that are serious enough to warrant healthcare resource utilisation. This is unlike anatomic and physiologic criteria, which are direct clinical manifestations of sustained injuries and, when present, would more likely require the attention of the trauma team and subsequent healthcare resource utilisation. Boyle et al. (9) proposed that mechanism-of-injury criteria, in the absence of anatomic and physiologic criteria, were not clinically significant for major trauma. A Scandinavian study proposed revised criteria for TTA that place emphasis on anatomic and physiologic criteria while modifying mechanism of injury criteria by removing motorcycle accidents, considerable deformation of the vehicle passenger compartment, and traffic accidents at speeds greater than 60 kilometres per hour (10).

Specific to level-two trauma centres, a hospital in India identified the modified shock index, defined as heart rate divided by mean arterial pressure, as a potential marker for predicting mortality in adult trauma patients; it was significantly better than heart rate, systolic or diastolic blood pressure, and shock index (heart rate divided by systolic blood pressure), making it an ideal physiologic criterion for TTA (11). The under-triage and over-triage proportions, based on need for immediate emergency interventions or operative procedures, at another level-two trauma centre in the United States of America changed from 1% (under-triage) and 79% (over-triage), figures similar to our study, to 4% and 12%, respectively, after adopting a simplified criterion comprising only four variables-hypotension, mental status, altered respirations, and penetrating truncal wound (12).

Instead of a single-tier trauma team that activates all members of the multidisciplinary team, a tiered trauma team may be more prudent for deployment of personnel. A centre in the Netherlands changed the trauma system from a single-tier to a two-tier system, and this reduced the proportion of over-triage (13). A two-tier system was similarly implemented as Code Red in the United Kingdom (14) and as Code Crimson in Australia (15), in which the first tier comprised the ED team attending to patients who met the mechanism-of-injury criteria, and the second tier comprised the trauma team attending to patients who met additional anatomic or physiologic criteria. Given the widespread adoption of Advanced Trauma Life Support, the initial assessment and management of patients, including those at high risk for severe injuries, can be competently and confidently performed by the ED team (16), with the trauma team ready to augment care as needed. Applying this two-tier approach to our study would reduce 418 TTA in a single-tier system that were activated based on mechanism criteria to 81 if the trauma team attended only to patients who fulfilled additional anatomic or physiologic criteria after assessment by the ED team, corresponding to an 80.6% reduction in multidisciplinary team activations.

TTA not only provides manpower in terms of quantity and expertise. From a process perspective, TTA also provides a system to expedite care from the ED to investigations (e.g., computed tomography), to interventions (e.g., interventional radiology procedures or operating-theatre interventions), and to admission to areas with higher acuity of care, such as intensive care or high-dependency units. Therefore, instead of focusing on staff deployment, another possible solution for a single-tier trauma team is to implement system-level protocols to expedite care for patients with serious injuries. For instance, at another hospital in Singapore, the Critical Haemorrhage to Operating Room Patient protocol, consisting of clearly defined roles, responsibilities, and tasks for healthcare professionals, was implemented to facilitate the escalation and mobilisation of resources upon recognition of a severely injured patient; all ten patients in this case series met the target of access to definitive intervention within 90 minutes, with an average time of 73 minutes from arrival at the ED to transfer to interventional radiology or the operating theatre (17).

Study Limitations

As this study was conducted at a single ED in a level-two trauma centre in Singapore, our recommendations are more relevant to EDs with settings similar to ours. For instance, trauma care is a continuum from prehospital to in-hospital care; Singapore's compact size and the high density of hospitals with trauma capabilities would result in shorter response and transport times by an EMS that uses predefined protocols and has access to online medical control when attending to trauma patients. The retrospective nature of this study meant that documentation could not be standardised, resulting in incomplete or missing information relevant to this study, such as the trauma team's perception of the appropriateness of TTA. We were also unable to determine the impact of TTA on

other patients in the ED or the hospital. Furthermore, the low frequency of TTA, coupled with a small number of patients with severe injuries requiring high-level resource utilisation, precluded further analysis and efforts to modify our current activation criteria.

Conclusion

TTA was an uncommon event at the ED of a level two trauma centre. To reduce the proportion of over-triage, we proposed revising the activation criteria and implementing a tiered trauma team or expedited care approach, since the overall injury severity of patients and their utilisation of healthcare resources were low. Future work should focus on the prevalence and spectrum of severe injuries, because these patients are the priority of the local trauma system, and any improvement efforts should be based on this.

Ethics

Ethics Committee Approval: This study was approved by Institutional Review Board of SingHealth, Singapore (CRIB reference: 2021/2417, date: 03.07.2021).

Informed Consent: Waiver of informed consent as it was a retrospective review of trauma registry data.

Footnotes

Authorship Contributions

Surgical and Medical Practices: S. J. L., H. A., W. J. D. T., K. V., S. H. C. L., C. M. L., J. H. P., Y. L. C. L., Concept: C. M. L., J. H. P., Design: C. M. L., J. H. P., Data Collection or Processing: S. J. L., H. A., S. H. C. L., Analysis or Interpretation: S. J. L., H. A., W. J. D. T., K. V., J. H. P., Y. L. C. L., Literature Search: S. J. L., H. A., W. J. D. T., K. V., J. H. P., Y. L. C. L., Writing: S. J. L., H. A., W. J. D. T., K. V., J. H. P., Y. L. C. L.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Trauma service: trauma team roles and responsibilities. (n.d.). Avaible from: https://www.rch.org.au/trauma-service/manual/Trauma_team_roles_and_responsibilities/
2. Tiel Groenestege-Kreb D, van Maarseveen O, Leenen L. Trauma team. Br J Anaesth. 2014;113:258-65. Epub 2014 Jun 30.
3. Committee on trauma American College of surgeons, editors. Resources for optimal care of the injured patient, 6th ed. chicago, IL: American College of Surgeons. 2014.
4. James MK, Schubl SD, Francois MP, Doughlin GK, Lee SW. Introduction of a pan-scan protocol for blunt trauma activations: what are the consequences? Am J Emerg Med. 2017;35:13-9. Epub 2016 Sep 22.
5. Newgard CD, Staudenmayer K, Hsia RY, Mann NC, Bulger EM, Holmes JF, et al. The cost of overtriage: more than one-third of low-risk injured patients were taken to major trauma centers. Health Aff (Millwood). 2013;32:1591-9.
6. Falcone RA Jr, Haas L, King E, Moody S, Crow J, Moss A, et al. A multicenter prospective analysis of pediatric trauma activation criteria routinely used in addition to the six criteria of the American College of Surgeons. J Trauma Acute Care Surg. 2012;73:377-84; discussion 384.
7. Uleberg O, Vinjevoll OP, Eriksson U, Aadahl P, Skogvoll E. Overtriage in trauma - what are the causes? Acta Anaesthesiol Scand. 2007;51:1178-83.
8. Stuke LE, Duchesne JC, Hunt JP, Marr AB, Meade PC, McSwain NE. Mechanism of injury is not a predictor of trauma center admission. Am Surg. 2013;79:1149-53.
9. Boyle MJ, Smith EC, Archer F. Is mechanism of injury alone a useful predictor of major trauma? Injury. 2008;39:986-92. Epub 2008 Jul 31.
10. Dehli T, Fredriksen K, Osbakk SA, Bartnes K. Evaluation of a university hospital trauma team activation protocol. Scand J Trauma Resusc Emerg Med. 2011;19:18.
11. Singh A, Ali S, Agarwal A, Srivastava RN. Correlation of shock index and modified shock index with the outcome of adult trauma patients: a prospective study of 9860 patients. N Am J Med Sci. 2014;6:450-2.
12. Lehmann R, Brounts L, Lesperance K, Eckert M, Casey L, Beekley A, et al. A simplified set of trauma triage criteria to safely reduce overtriage: a prospective study. Arch Surg. 2009;144:853-8.
13. Harmsen AK, Giannakopoulos G, Azijli K, Biesheuvel T, Geeraedts LG, Bloemers F. Optimization of trauma care: a two-tiered inhospital trauma team response system. Arch Trauma Res. 2017;6:15.
14. Jenkins P, Rogers J, Kehoe A, Smith JE. An evaluation of the use of a two-tiered trauma team activation system in a UK major trauma centre. Emerg Med J. 2015;32:364-7.
15. Davis T, Dinh M, Roncal S, Byrne C, Petchell J, Leonard E, et al. Prospective evaluation of a two-tiered trauma activation protocol in an Australian major trauma referral hospital. Injury. 2010;41:470-4. Epub 2010 Jan 21.
16. Bouillon B, Kanz KG, Lackner CK, Mutschler W, Sturm J. Die bedeutung des advanced trauma life support (ATLS) im schockraum [The importance of advanced trauma life support (ATLS) in the emergency room]. Unfallchirurg. 2004;107:844-50. German.
17. Kang ML, Goo JTT, Lee DJK. CHOP protocol: streamlining access to definitive intervention for major trauma victims. Singapore Med J. 2021;62:620-2. Epub 2020 Jul 30.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Editorial
DOI: 10.4274/eajem.galenos.2026.65481
Eurasian J Emerg Med. 2026;25: 199-200

Improving the First-attempt Success with Hyperangulated Videolaryngoscopy in Emergency Settings

Raghuraman M Sethuraman

Sree Balaji Medical College & Hospital, Bharath Institute of Higher Education and Research (BIHER), Clinic of Anesthesiology, Chennai, India

Keywords: Videolaryngoscope, hyperangulated videolaryngoscopy, difficult airway, emergency settings

Improving the First-attempt Success with Hyperangulated Videolaryngoscopy in Emergency Settings

Airway management is particularly challenging, especially in emergencies. It may require one or more special devices in certain cases. The videolaryngoscope is an important device for rapidly securing the airway. Based on the type of blades in a videolaryngoscope, it is broadly classified into two types: (1) standard-geometry videolaryngoscopy (SGVL) and (2) hyperangulated videolaryngoscopy (HAVL).

Many studies were published on this device, most in operating room settings and a few in emergency department settings. However, studies focusing exclusively on difficult airway scenarios are limited in both settings. Recently, a systematic review and meta-analysis on the comparative efficacy of SGVL versus HAVL was published. It concluded that the SGVL should be prioritized for emergency airway management, while careful operator training is required if an HAVL is used (1).

The HAVL was introduced with the main purpose of improving the laryngeal view in difficult cases, thereby leading to a higher first-attempt success rate. However, the main reason the previous meta-analyses, as well as the recent one (1), could not definitively conclude that HAVL is a better choice in difficult conditions is probably the limited number of studies focusing exclusively on difficult airways in these settings. Notably, a recent study that included only difficult airway conditions of diverse origins concluded that the laryngeal view and first-attempt success rate were superior with the HAVL compared with the SGVL (2). Importantly, all cases in that study were managed by the expert operators in the operating-room setting (2) and therefore cannot be extrapolated to all situations, as emphasized by von Hellmann et al. (1).

The main focus should be the conclusions of the meta-analysis by von Hellmann et al. (1): "The superior optical view afforded by HAVL did not translate into higher rates of successful first-pass intubation." The improved laryngeal view with a videolaryngoscope results from the camera's position, which can bypass anatomical structures that obstruct the view during direct laryngoscopy. However, this, per se, cannot facilitate passage of the tracheal tube. Hence, it is unsurprising that good laryngeal views did not result in a higher first attempt success rate. Moreover, the grading of the laryngeal view on videolaryngoscopy is controversial and beyond the scope of this Editorial. Indeed, clinicians may be better served by focusing on a slightly compromised view ("worse or partial view") to facilitate passage of the tracheal tube (3).

Furthermore, I agree with the recommendation of von Hellmann et al. (1) "The need for dedicated training, potentially including simulation and deliberate practice, before relying on HAVL for emergency airway management." Nevertheless, I would like to add a few important points in addition to noting the lack of expertise. 1. The proper "sniffing position" should be strictly adopted for difficult cases warranting an HAVL, except where it is contraindicated. This point is often overlooked in current practice. 2. It is advisable to use the specialized stylet offered by the company rather than other personal choices. This is especially

Corresponding Author: Raghuraman M Sethuraman MD, Sree Balaji Medical College & Hospital, Bharath Institute of Higher Education and Research (BIHER), Clinic of Anesthesiology, Chennai, India
E-mail: drraghuram70@gmail.com **ORCID ID:** orcid.org/0000-0001-8464-7458

Cite this article as: Sethuraman RM. Improving the first-attempt success with hyperangulated videolaryngoscopy in emergency settings. Eurasian J Emerg Med. 2026;25: 199-200.

Received: 25.11.2025
Accepted: 11.02.2026
Published: 27.02.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

true considering the HAVL's unique geometry and variations among manufacturers. Here, I would like to emphasize that it is better to stick to the stylet during the training phase, and the choice between stylet and bougie might be left to the individual preference for the experts. In rare cases, it is possible only to pass the introducer (bougie) under vision, and the tracheal tube subsequently over it. 3. The external laryngeal manipulation by an assistant, with close on-screen monitoring, would contribute further to the successful passage of the tracheal tube.

Conclusion

I suggest that future training focus on the three points as discussed above. Designing a randomized study presents many practical challenges in emergency settings. Therefore, I recommend conducting an observational, multicenter study involving a large number of participants, focusing exclusively on difficult airway scenarios in which HAVL is applied, and incorporating these three points.

Footnotes

Financial Disclosure: The author declared that this study received no financial support.

References

1. von Hellmann R, Dorn de Carvalho JM, Subramaniam A, Bellolio F, Gerberi D, Maia IWA, et al. Effect of hyperangulated videolaryngoscopy on first-attempt success in tracheal intubations: a systematic review and meta-analysis. Am J Emerg Med. 2025;98:191-200. Epub 2025 Aug 28.
2. Köhl V, Wünsch VA, Müller MC, Sasu PB, Dohrmann T, Peters T, et al. Hyperangulated vs. macintosh videolaryngoscopy in adults with anticipated difficult airway management: a randomised controlled trial. Anaesthesia. 2024;79:957-66. Epub 2024 May 24.
3. Ward PA, Karabulut G. Hyperangulated videolaryngoscopy: stylet, bougie or more practise? Anaesthesia. 2026;81:309-10. Epub 2025 Sep 14.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.91129
Eurasian J Emerg Med. 2026;25: 201-6

The Influence of Clinical Clerkship Rotation Sequence on Emergency Simulation Performance: A Cohort Study on Sixth-year Medical Students

Safinaz Alshiakh[1], Abdullah Algarni[2]

[1]King Abdulaziz University Faculty of Medicine, Department of Emergency Medicine, Jeddah, Saudi Arabia
[2]University of Jeddah Faculty of Medicine, Department of Medicine, Jeddah, Saudi Arabia

Abstract

Aim: The length and sequence of clinical clerkship rotations are essential factors that influence student learning outcomes and clinical performance. This study aimed to investigate whether clerkship sequences influence the performance of sixth-year medical students in emergency simulations of life-threatening scenarios.

Materials and Methods: This retrospective observational cohort study, conducted at our academic institution during the academic year 2024-2025, involved 334 sixth-year medical students divided into two groups based on initial clerkship general surgery (n=169) or general medicine (n=165) who switched at mid-year. Student performance was assessed by the emergency medicine faculty across five competency domains on a 0-5 scale during two high-fidelity emergency simulation scenarios: major trauma and acute myocardial infarction (AMI). Mann-Whitney U test was used to compare performances.

Results: Median scores across the performance domains clinical assessment, critical thinking, patient care, communication, and professional behavior were similar between the student groups in both simulations, with median scores ranging from 1 to 4. In the AMI scenario, surgery-first students had slightly higher overall median scores [13, interquartile range (IQR): 11-15] than medicine-first students (12, IQR: 10-15). However, the Mann-Whitney U test showed no statistically significant differences for either the AMI (p=0.120-0.86) or the trauma simulations (p=0.542-0.742).

Conclusion: Our study indicates that the order of clinical clerkship rotations did not significantly affect sixth-year medical students' performance in emergency simulations involving life-threatening cases. These findings inform ongoing discussions on optimizing medical education curricula. We recommend further research across multiple institutions to explore the effects of clerkship sequencing in greater detail.

Keywords: Clerkship, rotation sequence, clinical performance, emergency, medical students, high fidelity simulation, medical education

Introduction

During undergraduate medical education, clinical clerkships play a critical role in transforming medical students into skilled practitioners. These clerkships are complex curricular components shaped by numerous factors beyond the control of clerkship directors and medical education leadership, including patient demographics, healthcare team dynamics, curriculum duration, and evolving healthcare system policies (1). This complexity makes standardizing clerkship experiences and accurately assessing the impact of curriculum changes on student performance challenging (2). Throughout their medical education, from the early undergraduate years through the final clinical years (fifth and sixth years), students rotate through various clerkships, each providing diverse clinical exposure and skills. The sequence and duration of clerkships may enhance students' preparedness for clinical practice by influencing their clinical reasoning, procedural skills, and interdisciplinary collaborations.

Corresponding Author: Safinaz Alshiakh MD, King Abdulaziz University Faculty of Medicine, Department of Emergency Medicine, Jeddah, Saudi Arabia
E-mail: alshiakh@kau.edu.sa **ORCID ID:** orcid.org/0000-0003-3656-1048

Cite this article as: Alshiakh S, Algarni A. The influence of clinical clerkship rotation sequence on emergency simulation performance: a cohort study on sixth-year medical students. Eurasian J Emerg Med. 2026;25: 201-6 .

Received: 27.11.2025
Accepted: 11.02.2026
Published: 04.03.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Additionally, incorporating simulation-based education during rotations can be an effective teaching and assessment method to enhance students' competencies, knowledge, and clinical abilities (3-5). From an educational perspective, clerkship order may shape clinical competencies relevant to emergency care. Early exposure to general surgery can enhance procedural skills, rapid decision-making, and confidence in managing unstable or trauma patients (4). Conversely, early general medical training may strengthen diagnostic reasoning, clinical assessment, and systematic management of acute medical conditions. Differences in team structure, leadership roles, and levels of responsibility between the surgical and medical services may also influence students' communication and teamwork skills, which are key competencies assessed during emergency simulations.

Both clerkship length and sequence are considered critical factors influencing learning outcomes. The literature identifies various elements associated with student performance across clinical clerkships from the third year to the sixth year, including rotation order, student-reported stress, academic performance, and clinical evaluations (6,7). Indeed, the order of clerkship rotations has been shown to affect students' performance on standardized examinations, including those administered by the National Board of Medical Examiners (NBME) and the Clinical Subject Examinations of the United States Medical Licensing Examination (USMLE) (8). Despite this, the literature presents conflicting findings regarding the impact of clerkship length on examination performance. Several studies examining NBME scores across shortened clinical rotations in various specialties have reported either no effect or, in some cases, positive outcomes (9). One study identified the USMLE Step 1 scores as predictive of which students were at risk of poor performance on NBME subject exams (6). Strowd et al. (2) found no significant differences in student performance or satisfaction following curriculum shortening, suggesting that clerkship length may not be as critical as once assumed.

Although previous studies have addressed the importance of clerkship length, the influence of clerkship sequence on simulation practice performance, particularly during clinical years, remains underexplored. Research indicates that the timing and order of specific clerkships can influence academic outcomes, highlighting the importance of clerkship sequencing (10). Similarly, Hampton et al. (11) reported that rotation sequence could influence educational outcomes, supporting the idea that clerkship order may contribute to the development of clinical skills. Nevertheless, limited research has examined how the sequence of clinical rotations affects clinical performance during emergency simulations, a critical area of assessment for preparing future physicians.

This study investigated whether the order of clerkships for sixth-year medical students during their academic year affects their simulation performance, particularly in managing life-threatening scenarios. This investigation highlights the strategies for improving clinical competencies, medical education, and emergency readiness.

Materials and Methods

This retrospective observational study was conducted among sixth-year medical students at our academic institution during their emergency medicine (EM) rotation and was approved by the Biomedical Ethics Unit. The EM rotation, integrated into students' general surgery clerkship, was scheduled throughout the academic year. This study was approved by the biomedical Ethical Committee at King Abdulaziz University, Jeddah, Saudi Arabia (reference no: 395-24, date: 05.12 2024). A total of 334 students were divided into two groups: 169 who began with the general surgery rotation and the remaining 165 who began with the general medicine rotation. The groups switched midway through the academic year.

During the EM block, all students underwent routine assessments, including two high-fidelity simulation scenarios: a major trauma case and an acute myocardial infarction (AMI). These simulations were conducted at the University's Clinical Skills and Simulation Center using high-fidelity manikins. The standardized simulation process involved prebriefing, hands-on simulation, and debriefing sessions. Student performance was evaluated across five domains: clinical assessment and data gathering; critical thinking and decision-making; patient care; communication and teamwork; professional behavior. Each domain is scored from 0 to 5, with higher scores indicating better performance (Table 1).

Assessments were conducted by EM faculty members who were specifically trained in simulation assessments. They participated in standardization sessions to ensure consistent scoring. Data from students who completed both simulation scenarios with complete performance records were anonymized for analysis. This study aimed to compare students' performance based on their rotation preceding the EM block.

Statistical Analysis

Data analysis was performed using the Statistical Package for the Social Sciences, version 28 employing non-parametric descriptive and analytical methods. The median and interquartile range (IQR) were calculated to summarize descriptive statistics, while the Mann-Whitney U test was used for analytical comparisons. Statistical significance was defined as a p-value of less than 0.05.

Results

Our study compared the clinical performance of two groups of sixth-year medical students based on their rotation sequences during two high-fidelity simulation cases: AMI and major trauma. One group started with a general surgery clerkship (n=169), whereas the other began with a general medicine clerkship (n=165). The median scores across performance domains clinical assessment, critical thinking, patient care, communication, and professional behavior were generally similar between groups across both simulation cases, ranging from 1 to 4 (Tables 1, 2).

For the AMI simulation, the group that started with surgery demonstrated slightly higher overall performance, with a median total score of 13 (IQR: 11-15) versus 12 (IQR: 10-15) for the general medicine group; mean ranks were 175.59 and 159.22, respectively (Table 1). Despite these differences, the Mann-Whitney U test showed no statistically significant difference between the two groups across all domains (p=0.120-0.862). Similarly, in the major trauma simulation, performance scores were closely matched between groups, with a median score of 3 (IQR: 1-4) across most domains. The overall median performance

Table 1. Evaluation of the 6th year medical students for their clinical performance of a case of acute myocardial infarction, based on clerkship rotation sequence

Comparing items	Group started their year with a general surgery clerkship n=169 Median (IQR) Mean rank	Group started with general medicine and switched at mid-year n=165 Median (IQR) Mean rank	p-value*
Clinical assessment and data gathering	2 (1-4) 171.41	2 (1-4) 163.50	0.448
Critical thinking	3 (1-4) 173.24	2 (1-4) 161.62	0.265
Patient care	3 (1-4) 172.45	2 (1-4) 162.43	0.336
Communication and collaboration	2 (1-4) 165.09	3 (1-4) 169.97	0.640
Professional behaviour	2 (1-4) 166.60	3 (1-4) 168.42	0.862
Total	13 (11-15) 175.59	12 (10-15) 159.22	0.120

*Mann-Whitney U test, IQR: Interquartile range

Table 2. Evaluation of the 6th year medical students for their clinical performance of a case of major trauma, based on clerkship rotation sequence

Comparing item	Group started their year with a general surgery clerkship n=169 Median (IQR) Mean rank	Group started with general medicine and switched at mid-year n=165 Median (IQR) Mean rank	p-value*
Clinical assessment and data gathering	3 (1-4) 169.84	3 (1-4) 165.10	0.649
Critical thinking	2 (1-4) 170.75	2 (1-4) 164.17	0.527
Patient care	2 (1-4) 165.81	3 (1-4) 169.23	0.742
Communication and collaboration	3 (1-4) 164.34	3 (1-4) 170.73	0.540
Professional behaviour	3 (1-4) 170.15	2 (1-4) 164.79	0.607
Total	13 (10-15) 170.67	13 (10-15) 164.25	0.542

Mann-Whitney U test, IQR: Interquartile range

scores were 13 (IQR: 10-15) for both groups, and the mean ranks were nearly equivalent (170.67 vs. 164.25) (Table 2). The statistical analysis revealed no significant differences between the groups: all p-values exceeded 0.05 (range: 0.542-0.742).

These findings indicate that the sequence of initial clinical clerkship rotations did not significantly influence students' clinical performance during simulated emergency cases, suggesting that both groups demonstrated comparable competence regardless of their clerkship sequence.

Discussion

Evaluating the impact of clerkship rotations on medical students' learning and performance requires consistent assessment tools and procedures applied before and after the clerkship or through repeated measures to track progress over time (12). However, medical schools often vary in their structure, the sequencing of disciplinary clerkship rotations, and methods of student evaluation during clinically integrated years, making it difficult to standardize assessments (13). Clinical knowledge is inherently complex and multidimensional, presenting challenges in accurately measuring the learning growth derived from clerkship experiences. Additionally, it is logistically difficult to schedule all clerkship rotations simultaneously, complicating efforts to effectively assess the impact of clerkship programs on students' clinical knowledge development.

In this study, we compared medical students who started their year with a general surgery clerkship to those who began with general medicine and who switched midway through the year. No significant differences were observed between these groups in performance on the high-fidelity simulation cases involving major trauma and AMI (Tables 1, 2). The literature review did not reveal any directly comparable studies that corroborated our findings. However, Phares et al. (10) examined the effect of the timing of surgery and internal medicine clerkships on surgery shelf examination scores in the United States. They concluded that students who completed an internal medicine clerkship before a surgery clerkship achieved higher scores on the NBME Surgery Shelf examination. Additionally, their findings suggest that controlling for the timing of rotations yielded no significant differences between the groups compared. Similar conclusions have been drawn by other researchers, including Adelsheimer et al. (14) and Dong et al. (15), who reported no clear advantage of one clerkship sequence over another. Dong et al. (15) further investigated the impact of clerkship sequence on performance and found that completing internal medicine before surgery led to higher surgical subject examination scores, which, in turn, predicted subsequent USMLE Step scores and explained 28% of Step 2 CK variance. However, these findings are not directly comparable to those of our study, as we focused specifically on clinical performance during emergency simulation cases, rather than on standardized exams.

In our study, a comparison of the clinical performance between the two groups revealed slightly higher mean rank scores for students who started their year with a general surgery clerkship. However, the differences were not statistically significant, suggesting that clerkship sequence may not substantially influence clinical performance during high-fidelity simulation assessments. Although the overall performance was summarized using total simulation scores, domain-specific analyses were performed across the five competency areas. Trauma and AMI simulations differ in their cognitive and procedural demands, and the clerkship sequence could plausibly influence specific competencies, such as clinical assessment or communication. However, no statistically significant differences were observed across individual domains in either simulation scenario, suggesting broadly comparable competency development across clerkship sequences within the limits of the assessment tools. Prospective studies incorporating baseline academic metrics, prior simulation exposure, and multivariate modeling are needed to evaluate the independent effects of clerkship sequencing with greater precision. Furthermore, Gao et al. (8) conducted a study including medical students from three medical schools and found that starting the clerkship year in family medicine, internal medicine, pediatrics, or surgery did not significantly affect subsequent performance on shelf examinations or on the USMLE Step 2 Clinical Knowledge. This suggests that the sequence of clerkship rotations is not associated with improved clinical performance, higher NBME shelf scores, or higher Step 2 Clinical Knowledge examination scores.

In the broader context of evaluating clinical knowledge development, the Comprehensive Clinical Science Examination (CCSE) is a valuable formative assessment tool designed to provide students with feedback on their progress across various clinical rotations, such as internal medicine, surgery, and pediatrics (16). Administered by the NBME, the CCSE serves as a reliable and standardized tool for assessing competency-based medical education progress (17). By analyzing CCSE scores at different stages before, during, and after clinical clerkships researchers can gain valuable insights into how various rotations contribute to the development of students' disciplinary knowledge and clinical skills. Experiential learning theory, which emphasizes learning through experience and reflection, has been widely applied in clinical rotations and clerkships to enhance students' practical skills and knowledge acquisition (18). Chang et al. (16) used the CCSE and segmented regression analysis to assess the impact of clinical clerkships on medical students' disciplinary knowledge,

revealing significant improvements, particularly in obstetrics and gynecology and in psychiatry, with smaller gains in surgery. Their findings provided meaningful insights into curriculum design and assessment.

Study Limitations

This study had several limitations. First, the findings may not be generalizable because the study was conducted at a single college of medicine with a specific curriculum structure. Variations in clerkship design across institutions may influence performance. Second, the analyses were limited to univariate comparisons without adjusting for potential confounders. Therefore, unmeasured differences between the groups may account for the lack of statistically significant effects. Third, although a standardized 0-5 scale was used across the performance domains, the simulation tool was designed to assess competency attainment rather than detect subtle group differences. Despite a full score range and the absence of floor or ceiling effects, subtle educational influences of the clerkship sequence may not have been detected. Furthermore, our study lacked an equivalence framework and an a priori power calculation; therefore, non-significant findings reflect limited statistical power to detect differences and provide no evidence of equivalence between clerkship sequences.

Conclusion

Our study found no significant differences in emergency simulation performance between groups based on the clerkship rotation sequence. While existing literature suggests that clerkship timing can influence performance on standardized examinations, our findings do not support the notion that completing internal medicine before surgery enhances clinical performance in emergency cases. These results contribute to a broader discussion of optimizing medical education and may inform future research and curriculum development aimed at improving clinical training and assessments. Further studies at other universities are recommended to validate our findings and to provide deeper insights into the impact of clerkship rotation sequences.

Ethics

Ethics Committee Approval: This study was approved by the biomedical Ethical Committee at King Abdulaziz University, Jeddah, Saudi Arabia (reference no: 395-24, date: 05.12 2024).

Informed Consent: Informed consent was obtained from all participants involved in the study.

Footnotes

Authorship Contributions

Design: S.A., Data Collection or Processing: A.A., Analysis or Interpretation: S.A., Literature Search: A.A., Writing: S.A., A.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Kraakevik JA, Beck Dallaghan GL, Byerley JS, Monrad SU, Davis JA, Hammoud MM, et al. Managing expansions in medical students'clinical placements caused by curricular transformation: perspectives from four medical schools. Med Educ Online. 2021;26:1857322.
2. Strowd LC, Hartman N, Askew K, Vallevand A, McDonough K, Goforth J, et al. The impact of shortened clinical clerkships on medical student performance and clerkship assessment. Med Sci Educ. 2021;31:1333-41.
3. Chakravarthy B, Ter Haar E, Bhat SS, McCoy CE, Denmark TK, Lotfipour S. Simulation in medical school education: review for emergency medicine. West J Emerg Med. 2011;12:461-6.
4. McLaughlin S, Fitch MT, Goyal DG, Hayden E, Kauh CY. Simulation in graduate medical education 2008: a review for emergency medicine. Acad Emerg Med. 2008;15:1117-29.
5. Okuda Y, Bryson EO, DeMaria S Jr, Jacobson L, Quinones J, Shen B, et al. The utility of simulation in medical education: what is the evidence? Mt Sinai J Med. 2009;76:330-43.
6. Casey PM, Palmer BA, Thompson GB, Laack TA, Thomas MR, Hartz MF, et al. Predictors of medical school clerkship performance: a multispecialty longitudinal analysis of standardized examination scores and clinical assessments. BMC Med Educ. 2016;16:128.
7. Helou MA, Keiser V, Feldman M, Santen S, Cyrus JW, Ryan MS. Student well-being and the learning environment. Clin Teach. 2019;16:362-6.
8. Gao H, Askew K, Violato C, Manthey D, Burns C, Vallevand A. Does clerkship rotation sequence affect performance on National Board of Medical Examiners (NBME) clinical subject examinations and United States Medical Licensing Examination (USMLE) step 2 clinical knowledge (CK) examination? Med Sci Educ. 2019;29:763-70.
9. Fitz MM, Adams W, Haist SA, Hauer KE, Ross LP, Raff A, et al. Which internal medicine clerkship characteristics are associated with students' performance on the NBME medicine subject exam? A multi-institutional analysis. Acad Med. 2020;95:1404-10.
10. Phares A, Sauder CA, Salcedo ES, Leshikar DE, Irwin C, Middleton G, et al. Timing of surgery and internal medicine clerkships and surgery shelf examination scores. J Surg Res. 2019;244:456-9. Epub 2019 Jul 19.
11. Hampton HL, Collins BJ, Perry KG Jr, Meydrech EF, Wiser WL, Morrison JC. Order of rotation in third-year clerkships. Influence on academic performance. J Reprod Med. 1996;41:337-40.
12. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther. 2002;27:299-309.
13. Riese A, Rappaport L, Alverson B, Park S, Rockney RM. Clinical performance evaluations of third-year medical students and association with student and evaluator gender. Acad Med. 2017;92:835-40.

14. Adelsheimer A, Berman RS, Pachter HL, Hochberg MS. Surgical clerkship or medical clerkship first: does it make a difference? Am J Surg. 2018;215:304-8. Epub 2017 Nov 11.

15. Dong T, Copeland A, Gangidine M, Schreiber-Gregory D, Ritter EM, Durning SJ. Factors associated with surgery clerkship performance and subsequent USMLE step scores. J Surg Educ. 2018;75:1200-5. Epub 2018 Mar 12.

16. Chang C, Laird-Fick HS, Mitchell JD, Parker C, Solomon D. Assessing the impact of clerkships on the growth of clinical knowledge. Ann Med. 2025;57:2443812. Epub 2024 Dec 28.

17. National Board of Medical Examiners. Comprehensive Clinical Science Examination [CCSE]; 2024. Avaible from: https://www.nbme.org/subject-exams/comprehensive/clinical-science?utm_source=chatgpt.com

18. Kolb DA, Boyatzis RE, Mainemelis C. Experiential learning theory: previous research and new directions. In: perspectives on thinking, learning, and cognitive styles. New York, NY: Routledge. 2014:227-48.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Brief Report
DOI: 10.4274/eajem.galenos.2026.65625
Eurasian J Emerg Med. 2026;25: 207-9

Multimodal AI-assisted Diagnosis and Management of Cardiac Emergencies: A Pilot Study

Evren Ekingen[1], Mete Üçdal[2]

[1]Etimesgut Şehit Sait Ertürk State Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[2]Etimesgut Şehit Sait Ertürk State Hospital, Clinic of Internal Medicine, Ankara, Türkiye

Abstract

To evaluate the diagnostic accuracy and guideline adherence of a multimodal artificial intelligence (AI) system integrating specialized models for electrocardiogram (ECG) interpretation, clinical reasoning, and risk stratification in acute cardiac emergencies. We retrospectively analyzed 32 consecutive patients who presented to Etimesgut Şehit Sait Ertürk State Hospital, Ankara, Türkiye, with cardiac emergencies between September 2024 and January 2025. A multimodal AI system integrating GPT-4o with vision capabilities for ECG interpretation, Claude 3.5 Sonnet for clinical reasoning, and Gemini 2.0 Flash for risk stratification was compared with standard clinical practice. The cohort comprised patients with acute coronary syndrome (ACS) (n=14), acute pericarditis (n=6), pulmonary embolism (n=5), aortic dissection (n=4), and myocarditis (n=3). Diagnostic accuracy, guideline concordance, and clinical outcomes were assessed. The multimodal AI achieved an overall diagnostic accuracy of 93.8% (30/32) and correctly classified all ACS subtypes (14/14, 100%), including ST-elevation myocardial infarction (STEMI) localization. Pericarditis, pulmonary embolism, aortic dissection, and myocarditis were correctly identified in 83.3% (5/6), 80.0% (4/5), 75.0% (3/4), and 66.7% (2/3) of cases, respectively. Differentiation between STEMI and pericarditis based on ST-segment morphology achieved an accuracy of 91.2%. Guideline adherence was 96.9% (31/32) for the multimodal AI versus 84.4% (27/32) for clinical practice (p=0.046). The AI correctly identified all high-risk patients requiring intensive monitoring. Multimodal AI systems integrating specialized models for ECG interpretation, clinical reasoning, and risk stratification demonstrate high diagnostic accuracy and guideline adherence in cardiac emergencies, approaching human expert-level performance. These pilot results support larger prospective validation trials examining multimodal AI integration in time-critical cardiovascular care.

Keywords: Artificial intelligence, cardiac emergencies, large language models, multimodal AI, electrocardiography, clinical decision support

Dear Editor,

The integration of multimodal artificial intelligence (AI) systems into emergency cardiovascular care represents a paradigm shift in clinical decision support (1,2). Recent advances in vision-language models have enabled direct electrocardiogram (ECG) image interpretation combined with sophisticated clinical reasoning, potentially addressing diagnostic challenges in cardiac emergencies with overlapping presentations (3). We present a pilot evaluation of a multimodal AI system across the spectrum of acute cardiac conditions.

We retrospectively analyzed 32 consecutive patients presenting with cardiac emergencies to Etimesgut Şehit Sait Ertürk State Hospital, Ankara, Türkiye, between September 2024 and January 2025. A multimodal AI system integrating GPT-4o with vision capabilities for ECG interpretation, Claude 3.5 Sonnet for clinical reasoning, and Gemini 2.0 Flash for risk stratification was compared against standard clinical practice. The cohort (mean age 57.4±15.8 years; 71.9% male) comprised acute coronary syndrome (ACS; n=14), acute pericarditis (n=6), pulmonary embolism (PE; n=5), aortic dissection (n=4), and myocarditis (n=3).

The multimodal AI achieved 93.8% overall diagnostic accuracy (30/32), correctly classifying all ACS subtypes (14/14, 100%), including localization of ST-elevation myocardial infarction

Corresponding Author: Mete Üçdal MD, Etimesgut Şehit Sait Ertürk State Hospital, Clinic of Internal Medicine, Ankara, Türkiye
E-mail: meteucdal@hacettepe.edu.tr **ORCID ID:** orcid.org/0000-0002-2321-2642

Cite this article as: Ekingen E, Üçdal M. Multimodal AI-assisted diagnosis and management of cardiac emergencies: a pilot study. Eurasian J Emerg Med. 2026;25: 207-9

Received: 09.01.2026
Accepted: 27.02.2026
Published: 05.03.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

(STEMI), and demonstrating particular strength in differentiating STEMI from pericarditis based on ST-segment morphology (concave vs. saddle-shaped elevation, PR depression, reciprocal changes). Pericarditis was correctly identified in 5/6 cases (83.3%), with one case initially misclassified as inferior STEMI due to prominent ST-elevation. PE diagnosis achieved an 80.0% sensitivity (4/5), while aortic dissection and myocarditis showed sensitivities of 75.0% (3/4) and 66.7% (2/3), respectively (Table 1).

Guideline-adherence analysis revealed 96.9% concordance (31/32) with current European Society of Cardiology/American College of Cardiology recommendations for the multimodal AI system, compared with 84.4% (27/32) for clinical practice (p=0.046). The AI correctly recommended primary percutaneous coronary intervention for all STEMI cases, a colchicine-aspirin combination for pericarditis based on COPE-3 trial evidence, and risk-stratified anticoagulation for PE using simplified PESI scores. Clinical outcomes showed one in-hospital death (aortic dissection, 3.1%) and major complications in 3 patients (9.4%). The AI prospectively identified all high-risk patients requiring intensive monitoring (Table 2).

Our findings align with the rapidly evolving landscape of multimodal AI in cardiovascular medicine. Lee et al. (4) recently compared ChatGPT and Gemini with dedicated ECG AI tools for myocardial infarction detection, reporting that specialized algorithms significantly outperformed general-purpose large language models, which supports our multi-model integration strategy. Hilgendorf et al. (5) demonstrated in JACC advances that multimodal deep learning incorporating ECG signals with clinical variables achieved superior acute myocardial infarction detection compared to unimodal approaches, consistent with our 93.8% diagnostic accuracy. The TCT 2025 presentation by Herman et al. (6) showed that AI-based ECG analysis significantly improved STEMI detection while reducing false activations, highlighting the clinical readiness of these technologies. Furthermore, Yang et al. (7) proposed an integrated AI-driven cardiovascular platform emphasizing the paradigm shift from single diagnostic tools to comprehensive intelligent systems, which our multimodal architecture exemplifies. Our 91.2% success rate in differentiating pericarditis from STEMI addresses a critical clinical challenge, as overlapping ST-elevation patterns frequently lead to misdiagnosis in emergency settings. Limitations include a small sample size, a single-center design, and a retrospective methodology. Nevertheless, this pilot study suggests that multimodal AI systems warrant larger prospective validation trials, particularly to assess their ability to synthesize visual ECG data with clinical context in time-critical cardiac emergencies.

Table 1. Diagnostic performance of multimodal AI across cardiac emergency subtypes

Condition	n (%)	AI correct, n/N	Sensitivity, %	Specificity, %	PPV, %
ACS (STEMI/NSTEMI/UA)	14 (43.8)	14/14	100	94.4	93.3
Acute pericarditis	6 (18.8)	5/6	83.3	100	100
Pulmonary embolism	5 (15.6)	4/5	80.0	96.3	80.0
Aortic dissection	4 (12.5)	3/4	75.0	100	100
Myocarditis	3 (9.4)	2/3	66.7	100	100
Overall	**32 (100)**	**30/32**	**93.8**	**97.5**	**93.8**

ACS: Acute coronary syndrome, STEMI: ST-elevation myocardial infarction, NSTEMI: Non-ST-elevation myocardial infarction, UA: Unstable angina, PPV: Positive predictive value, AI: Artificial intelligence

Table 2. Guideline adherence and clinical outcomes

Parameter	Multimodal AI	Clinical practice
Overall guideline adherence, n/N (%)	31/32 (96.9)	27/32 (84.4)*
ACS: appropriate reperfusion strategy	14/14 (100)	12/14 (85.7)
Pericarditis: colchicine recommendation	6/6 (100)	4/6 (66.7)
PE: risk-stratified anticoagulation	5/5 (100)	5/5 (100)
High-risk patient identification, n/N (%)	4/4 (100)	3/4 (75.0)
In-hospital mortality, n (%)	—	1 (3.1)
Major complications, n (%)	—	3 (9.4)

*p=0.046 vs. Multimodal AI. ACS: Acute coronary syndrome, PE: Pulmonary embolism, AI: Artificial intelligence

Ethics

Ethics Committee Approval: Approved by Ankara Provincial Health Directorate Ethics Committee (decision no: 2025-10-3, date: 10.03.2025).

Informed Consent: This is a study involving a single-centered design and a retrospective methodology.

Footnotes

Authorship Contributions

Concept: E.E., M.Ü., Design: E.E., M.Ü., Data Collection or Processing: E.E., M.Ü., Analysis or Interpretation: E.E., M.Ü., Literature Search: E.E., M.Ü., Writing: E.E., M.Ü.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Rao SV, O'Donoghue ML, Ruel M, Rab T, Tamis-Holland JE, Alexander JH, et al. 2025 ACC/AHA/ACEP/NAEMSP/SCAI guideline for the management of patients with acute coronary syndromes: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 2025;151:e771-e862. Epub 2025 Feb 27. Erratum in: Circulation. 2025;151:e865. Erratum in: Circulation. 2025;151:e1098. Erratum in: Circulation. 2025;152:e402.
2. Wells BJ, Nguyen HM, McWilliams A, Pallini M, Bovi A, Kuzma A, et al. A practical framework for appropriate implementation and review of artificial intelligence (FAIR-AI) in healthcare. NPJ Digit Med. 2025;8:514.
3. Moreno-Sánchez PA, Del Ser J, van Gils M, Hernesniemi J. A design framework for operationalizing trustworthy artificial intelligence in healthcare: requiremen, tradeoffs and challenges for its clinical adoption. Information Fusion. 2026;127:103812.
4. Lee H, Yoo S, Kim J, Cho Y, Suh D, Lee K. Comparative diagnostic performance of a multimodal large language model versus a dedicated electrocardiogram AI in detecting myocardial infarction from electrocardiogram images: comparative study. JMIR AI. 2025;4:e75910.
5. Hilgendorf L, Petursson P, Andersson E, Rawshani A, Bhatt DL, Råmunddal T, et al. Fully automated diagnosis of acute myocardial infarction using electrocardiograms and multimodal deep learning. JACC Adv. 2025;4:102011.
6. Herman R, Mumma BE, Hoyne JD, Cooper BL, Johnson NP, Kisova T, et al. AI-enabled ECG analysis improves diagnostic accuracy and reduces false STEMI activations: a multicenter U.S. registry. JACC Cardiovasc Interv. 2026;19:145-56. Epub 2025 Oct 28.
7. Yang XY, Li YM, Wang JY, Jia YH, Yi Z, Chen M. Utilizing multimodal artificial intelligence to advance cardiovascular diseases. Precis Clin Med. 2025;8:pbaf016.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.78889
Eurasian J Emerg Med. 2026;25: 210-21

Evaluation of Sternum Fractures: A Single-institution Study

Salih Kür[1], Kadir Baturhan Çiflik[1], Ekin Zorlu[2], Nesimi Günal[3]

[1]Republic of Türkiye Ministry of Health, Kırıkkale Yüksek İhtisas Hospital, Clinic of Thoracic Surgery, Kırıkkale, Türkiye
[2]University of Health Sciences Türkiye, İzmir City Hospital, Clinic of Thoracic Surgery, İzmir, Türkiye
[3]Kırıkkale University Faculty of Medicine, Department of Thoracic Surgery, Kırıkkale, Türkiye

Abstract

Aim: Sternal fractures are relatively uncommon. Although most sternal fractures are managed conservatively, the presence of associated injuries to adjacent organs is associated with a significant increase in morbidity and mortality. This study aimed to evaluate the clinical characteristics, associated thoracic and extrathoracic injuries, management strategies, and outcomes of patients with traumatic sternal fractures.

Materials and Methods: Patients diagnosed with sternal fractures following thoracic trauma between January 2013 and May 2021 were retrospectively analyzed. Demographic data, mechanism of injury, fracture localization and type, concomitant injuries, seatbelt use, Injury Severity score (ISS), length of hospital stay, and mortality were assessed.

Results: Sternal fractures were identified in 230 of 2131 patients with thoracic trauma (10.7%). Traffic accidents were the most common mechanism of injury (n=189, 82.2%). Most fractures involved the sternal body (n=129, 56.1%) and were non-displaced (n=128, 55.7%). Retrosternal hematoma was detected in 10% of patients and was significantly associated with hemothorax and pulmonary contusion (p=0.021, p=0.033). Patients not using seatbelts had significantly higher ISS values and longer hospital stays (p=0.005, p=0.001). Thirteen patients (5.7%) died, predominantly due to severe associated injuries.

Conclusion: Although most sternal fractures can be managed conservatively, the presence of a retrosternal hematoma may indicate underlying pulmonary injury rather than cardiac involvement. Careful radiological evaluation and close in-hospital monitoring are essential to optimize outcomes and reduce mortality in these patients.

Keywords: Retrosternal hematoma, seatbelt, sternum fracture, thoracic trauma

Introduction

Trauma is a prevalent etiological risk of mortality throughout the first 40 years of life (1). Approximately 25% of these injuries are thoracic injuries (2). Thoracic trauma can be classified into two primary categories: blunt and penetrating. Blunt trauma constitutes the majority, accounting for approximately 75%, of thoracic injuries (3,4). Blunt thoracic trauma generally arises from motor vehicle collisions, falls from height, and direct impacts, whereas sternal fractures are relatively uncommon within this spectrum. Nonetheless, motor vehicle collisions represent the most frequent mechanism associated with sternal fractures (5-9).

The prevalence of sternum fractures varies according to age, gender, and the nature of the accident. Advanced age, female sex, and front-seat occupancy are established risk factors for sternal fractures (10). The primary determinant affecting morbidity in sternum fractures is associated injury to adjacent organs. Clinically, substantial morbidity and elevated mortality rates are evident, especially in cases of displaced fractures, mediastinal bleeding, cardiac contusion, and vascular injuries, which can increase mortality by 25-40% (6,10-12).

Conservative management can effectively address the majority of sternal fractures. However, in cases of displacement at the fracture site, significant discomfort, persistent non-union, and sternal instability, surgical fixation may be necessary (13,14). Timely diagnosis and intervention in these patients can enhance therapeutic results. Moreover, whereas approximately 95% of sternal fractures are successfully managed conservatively, the

Corresponding Author: Kadir Baturhan Çiflik MD, Republic of Türkiye Ministry of Health, Kırıkkale Yüksek İhtisas Hospital, Clinic of Thoracic Surgery, Kırıkkale, Türkiye
E-mail: baturhanciflik@gmail.com **ORCID ID:** orcid.org/0000-0002-5964-3778

Cite this article as: Kür S, Çiflik KB, Zorlu E, Günal N. Evaluation of sternum fractures: a single-institution study. Eurasian J Emerg Med. 2026;25: 210-21.

Received: 15.10.2025
Accepted: 03.03.2026
Published: 12.03.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

existence of underlying cardiac abnormalities necessitates prompt surgery, underscoring the critical importance of early detection (10,15).

The primary aim of this study is to provide a comprehensive analysis of patients with traumatic sternal fractures evaluated at a single center. The primary outcome is to assess the impact of concomitant thoracic and extrathoracic injuries on clinical outcomes. Secondary outcomes include an evaluation of demographic characteristics, trauma mechanisms, fracture localization and type, and the relationship between applied treatment strategies and the clinical course.

Materials and Methods

The study encompassed patients who presented to the emergency department with thoracic trauma between January 2013 to May 2021. The study was conducted as a retrospective cohort study including patients diagnosed with traumatic sternal fractures at a single center. Ethical approval was obtained from Ethics Committee Approval for Non-Interventional Research from Kırıkkale University on (decision number: 2021.07.02, date: 08.07.2021). The study was conducted in accordance with the Declaration of Helsinki established by the World Medical Association. The study excluded patients with non-traumatic sternal fractures, those with missing data, and patients whose thoracic computed tomography (CT) scans were inadequate for evaluating sternal fractures due to artifacts. The methodological framework of the study was structured according to the PICO (population): patients with traumatic sternal fractures; exposure: presence of retrosternal hematoma; comparator: patients without retrosternal hematoma; outcomes: mortality, pulmonary complications, and clinically relevant outcomes during hospitalization approach to clearly define the study design and analytical focus.

Clinical data were retrieved from the hospital information system, encompassing patients' age, sex, radiological imaging, laboratory results, mechanism of trauma, seatbelt usage in traffic accidents (based on patient self-report), length of hospitalization, Injury Severity score (ISS), concomitant thoracic and extra-thoracic injuries, fracture localization, and fracture type. Trauma severity was assessed using the ISS. A score below 16 was classified as low ISS, while a score of 16 or above was classified as high ISS. All thoracic CT scans were evaluated by a single experienced thoracic surgeon. Thoracic CT scans (64-slice multidetector CT, performed with intravenous (IV) contrast in arterial and/or venous phases, using 0.5-1-mm thin-section acquisition with multiplanar reconstructions) were evaluated as part of the routine trauma protocol. Thoracic CT was performed immediately after admission to the emergency department, following the initial assessment and resuscitation. Contrast-enhanced CT was preferred in cases of high-energy trauma and when physical examination findings suggestive of intrathoracic bleeding such as asymmetric respiration, hypoxia, subcutaneous emphysema, or crepitation were present. Additionally, contrast-enhanced thoracic CT was obtained in patients who demonstrated deterioration in vital signs during clinical follow-up and in those with suspected intrathoracic bleeding on radiologic surveillance.

Fracture localization was categorized according to the anatomical segments of the sternum: manubrium, corpus, xiphoid, and mixed regions. Fracture patterns were classified as displaced (a fracture characterized by disruption of cortical alignment between fracture fragments, with axial or translational displacement, angulation, or overlap between the fragments) or non-displaced (a fracture in which a fracture line is visible in the rib cortex, while anatomic alignment is preserved without significant displacement, angulation, or shortening between the fracture fragments) (16,17). A retrosternal hematoma was defined as a blood collection located in the anterior mediastinal space posterior to the sternum, secondary to trauma, appearing on non-contrast CT as a soft-tissue–density lesion [40-70 hounsfield unit (HU)] or a heterogeneous hyperdense area (18). A pulmonary contusion was defined as parenchymal lung injury representing alveolar hemorrhage and edema, characterized on CT by patchy ground-glass opacities or consolidations, typically adjacent to the pleura and not confined to segmental boundaries (19). A hemothorax was defined as the presence of pleural fluid with increased attenuation values (>35 HU) on CT, consistent with blood within the pleural space (20). Pneumomediastinum was defined as the presence of free air within the mediastinal space, visualized on CT as linear or focal air densities surrounding the trachea, esophagus, and major vessels (21).

All patients had chest CT for diagnostic purposes. Electrocardiograms were conducted on all patients with sternal fractures, and troponin I levels were assessed in individuals exhibiting abnormal electrocardiographic findings or those suspected of cardiac damage based on radiological imaging. Transthoracic echocardiography was performed in patients with increased troponin I levels (0-0.014 negative, 0.014-0.025 equivocal, 0.025-<10 positive). Additional diagnostic evaluations were initiated at the discretion of the thoracic surgeon, while a cardiologist performed echocardiography. The presence of myocardial contusion or pericardial effusion was considered diagnostic for blunt heart damage.

Surgical indications for sternal fractures included flail chest, displaced sternal fractures, open fractures, sternal fracture ends impinging on mediastinal organs, and the development of organ damage as surgical indications for sternal fractures.

The conservative treatment regimen comprised IV analgesics, suitable IV fluid resuscitation, nasal oxygen supplementation, and structured respiratory physiotherapy.

Statistical Analysis

Data were analyzed using IBM SPSS Statistics Standard Concurrent User, Version 26 (IBM Corp., Armonk, New York, USA). As the data demonstrated a normal distribution, descriptive statistics were expressed as mean ± standard deviation. The normality of numerical variables was assessed using the Shapiro-Wilk test, and homogeneity of variances was evaluated with the Levene's test. Since the numerical variables were normally distributed, comparisons with binary categorical variables were performed using the independent two-sample t-test. Comparisons between categorical variables were conducted using the chi-square test based on the Exact method. When the chi-square test yielded statistically significant results, subgroup analyses were performed using the two-proportion z-test with Bonferroni correction. A p-value<0.05 was considered statistically significant in all analyses.

Results

General Information About the Patients

The analysis demonstrated that sternum fractures occurred in 230 of 2131 patients (10.7%) who presented to the emergency department with thoracic injuries. Among these patients, 85 (37%) were female, while 145 (63%) were male. The mean age was 48.4±17.1 years. The predominant cause of injuries was traffic accidents (n=189, 82.2%). The mean length of hospitalization was 5.5±12.1 days. Thirteen individuals (5.7%) died (Table 1, Figure 1).

The predominant site of sternum fracture in our study was corpus (n=129, 56.1%). The majority of sternum fractures were non-displaced (n=128, 55.7%). Forty-nine patients (21.3%) sustained an isolated sternal fracture. The most common associated thoracic injury was rib fractures (n=140, 54.3%) while the most frequent accompanying-extra-thoracic injury was vertebral fractures (n=71, 36.2%). No patient underwent surgical sternal fixation due to the sternum fracture. Retrosternal hematoma was detected 23 patients (10%) with sternal fractures. All patients underwent chest CT for diagnostic evaluation. A total of 80 patients (34.8%) were suspected of having cardiac injury and therefore underwent troponin I measurement and echocardiographic evaluation. Echocardiography revealed no evidence of cardiac injury in any patient. A significant positive correlation was observed between troponin I levels and ISS (p<0.010).

No statistically significant association was observed between fracture localization and mortality (p=0.778). Among the patients who died, the predominant fracture location was the corpus (n=7). Concomitant rib fractures and intracranial hemorrhage were present-in ten mortality cases (76.9%). No mortality was observed among patients with isolated sternal fractures who were using seatbelts. All patients were managed conservatively, except for one patient with a retrosternal hematoma who developed hemothorax and subsequently underwent video-assisted thoracoscopic surgery (VATS) for hematoma evacuation.

Presence of Retrosternal Hematoma

Retrosternal hematoma occurred more frequently in patients with corpus fractures and displaced fractures, with statistically significant difference (p<0.001, p=0.001). Furthermore, a significant association was identified between retrosternal hematoma and the presence of hemothorax and pulmonary contusion (p=0.021, p=0.033). No statistically significant association was observed between the presence of retrosternal hematoma and ISS groups, ISS, length of hospitalization, or mortality (p=0.210, p=0.206, p=0.243, p=0.128) (Table 2).

Sternum Fracture and ISS Groups

The incidence of traffic accidents was higher in the low ISS group, and this difference was statistically significant (p=0.007). No statistically significant association was observed between fracture localization or fracture displacement and ISS groups (p=0.713, p=0.381). All isolated sternum fractures occurred in the low ISS group, which was statistically significant (p=0.002).

The low ISS group demonstrated a significantly higher incidence of rib fractures, hemothorax, pulmonary contusions, pneumothorax, facial injuries, scapular fractures, upper extremity fractures, and intra-abdominal organ injuries (p=0.005, p<0.001, p=0.046, p<0.001, p=0.001, p=0.008, p=0.023, p=0.032, p=0.016). In contrast, the high ISS group showed a significantly higher incidence of intracranial hemorrhage and lower extremity fractures (p<0.001, p=0.001). Both the length of hospitalization and mortality rate were significantly higher in the high ISS group (p=0.048, p<0.001) (Table 3).

Consequences of Seatbelt Use

Among patients using seatbelts, the predominant fracture localization was the corpus, and this difference was statistically significant (p=0.008). However, no statistically significant association was observed between fracture displacement and seatbelt usage (p=0.434). Vertebral fractures were more frequently observed in patients not using seatbelts, which was statistically significant (p=0.030). Patients without seatbelt use had significantly higher ISS values and longer hospitalization durations (p=0.005, p=0.001) (Table 4).

Table 1. General information of the patients	
	n (%)
Gender	
Male	145 (63)
Female	85 (37)
Age, (year)	
$\bar{x} \pm SD$	48.4±17.1
Troponin I, (ng/L)	
$\bar{x} \pm SD$	0.02±0.06
Trauma	
Vehicular traffic accident	189 (82.2)
Non-vehicle traffic accident	24 (10.4)
Fall from height	12 (5.2)
Animal kick	4 (1.7)
Assault	1 (0.4)
Fracture localization	
Manubrium	84 (36.5)
Sternal body	129 (56.1)
Xiphoid process	6 (2.6)
Mixed	11 (4.8)
Degree of fracture separation	
Non-displaced	128 (55.7)
Displaced	102 (44.3)
Retrosternal hematoma	
Yes	23 (10)
No	207 (90)
Isolated sternal fracture	
Yes	49 (21.3)
No	181 (78.7)
Seat belt use	
Yes	64 (33.9)
No	125 (66.1)
Comorbid thoracic injuries	
Rib fracture	140 (54.3)
Pulmonary contusion	50 (19.4)
Pneumothorax	42 (16.3)
Hemothorax	24 (9.3)
Pneumomediastinum	2 (0.7)
Comorbid extra-thoracic injuries	
Vertebral fracture	71 (36.9)
Intracranial injury	22 (11.5)
Pelvic fracture	17 (8.9)
Intra-abdominal organ injury	17 (8.9)
Clavicle fracture	15 (7.8)
Scapular fracture	15 (7.8)
Maxillofacial trauma	15 (7.8)
Upper extremity fracture	10 (5.2)
Lower extremity fracture	10 (5.2)
ISS group	
<16 (low)	199 (86.5)
≥16 (high)	31 (13.5)
ISS	
$\bar{x} \pm SD$	10.5±14.5
Length of hospitalization, (days)	
$\bar{x} \pm SD$	5.5±12.1
Mortality	13 (5.7)

n: Number of patients, %: Percentage of patients, $\bar{x}$: Mean, SD: Standard deviation, ISS: Injury Severity score

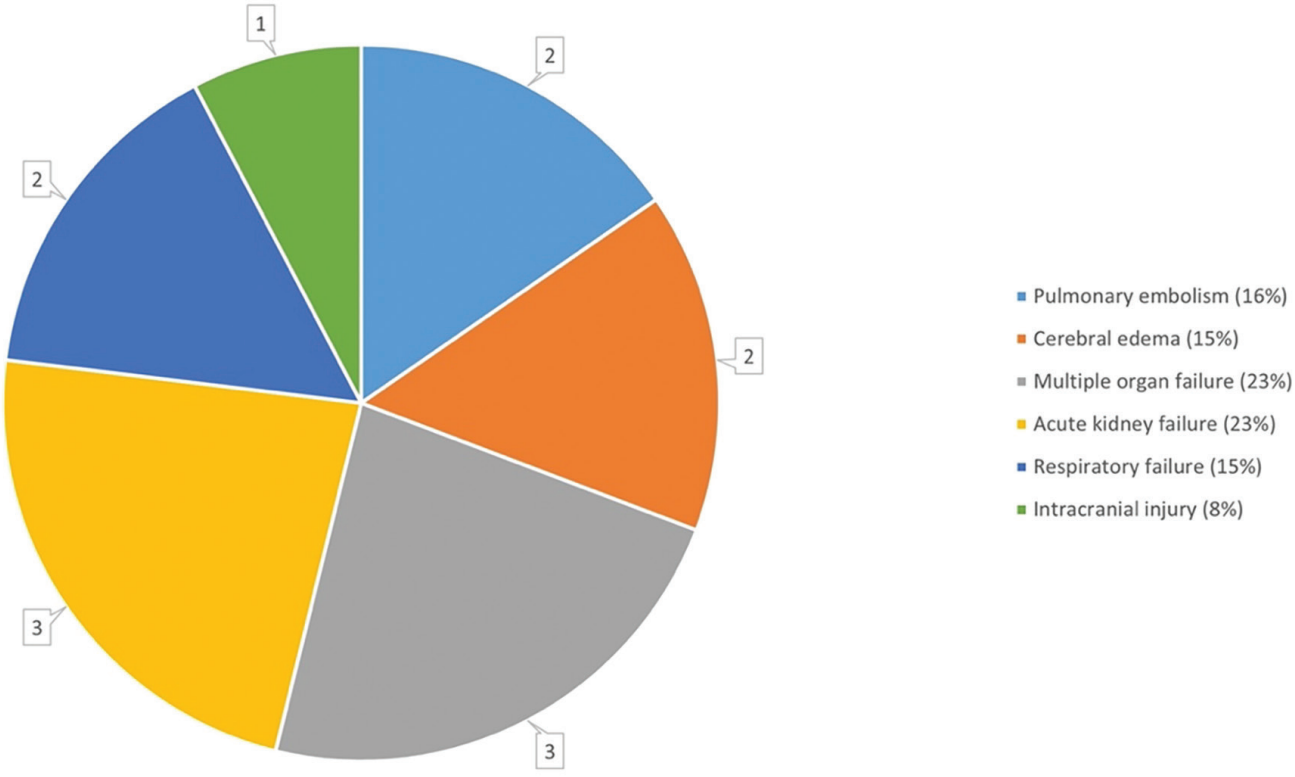


Figure 1. Causes of death

Table 2. Effects of the presence of retrosternal hematoma			
	Retrosternal hematoma present	**Retrosternal hematoma not present**	**p**
	n=23	**n=207**	
Gender Male Female	 8 (34.8) 15 (65.2)	 77 (37.2) 130 (62.8)	0.820
Age, (year) $\bar{x}$ ± SD	 54.4±15.8	 47.7±17.1	0.075
Troponin, (ng/L) $\bar{x}$ ± SD	 0.02±0.04	 0.02±0.06	0.851
Trauma Vehicular traffic accident Non-vehicle traffic accident Assault Animal kick	 19 (82.6) 3 (13) 0 (0) 0 (0)	 170 (82.1) 21 (10.1) 1 (0.5) 4 (1.9)	0.937
Fracture localization Manubrium Sternal body Xiphoid process Mixed	 7 (30.4) 9 (39.1) 0 (0) 7 (30.4)	 77 (37.2) 120 (58) 6 (2.9) 4 (1.9)	**<0.001**
Degree of fracture separation Non-displaced Displaced	 5 (21.7) 18 (78.3)	 123 (59.4) 84 (40.6)	0.001
Isolated sternal fracture Yes No	 5 (21.7) 18 (78.3)	 44 (21.3) 163 (78.7)	>0.999
Seat belt use Yes No	 7 (36.8) 12 (63.2)	 57 (33.5) 113 (66.5)	0.772

Table 2. Continued

	Retrosternal hematoma present	Retrosternal hematoma not present	p
	n=23	n=207	
Rib fracture Yes No	 11 (47.8) 12 (52.2)	 129 (62.3) 78 (37.7)	0.177
Hemothorax Yes No	 6 (26.1) 17 (73.9)	 18 (8.7) 189 (91.3)	0.021
Pulmonary contusion Yes No	 1 (4.3) 22 (95.7)	 49 (23.7) 158 (76.3)	0.033
Pneumothorax Yes No	 4 (17.4) 19 (82.6)	 38 (18.4) 169 (81.6)	>0.999
Pneumomediastinum Yes No	 1 (4.3) 22 (95.7)	 1 (0.5) 206 (99.5)	0.190
Intracranial injury Yes No	 4 (17.4) 19 (82.6)	 18 (8.7) 189 (91.3)	0.250
Maxillofacial trauma Yes No	 1 (4.3) 22 (95.7)	 14 (6.8) 193 (93.2)	>0.999
Clavicle fracture Yes No	 0 (0) 23 (100)	 15 (7.2) 192 (92.8)	0.374
Scapular fracture Yes No	 1 (4.3) 22 (95.7)	 14 (6.8) 193 (93.2)	>0.999
Vertebral fracture Yes No	 11 (47.8) 12 (52.2)	 60 (29) 147 (71)	0.093
Upper extremity fracture Yes No	 1 (4.3) 22 (95.7)	 9 (4.3) 198 (95.7)	>0.999
Pelvic fracture Yes No	 3 (13) 20 (87)	 14 (6.8) 193 (93.2)	0.389
Intra-abdominal organ injury Yes No	 2 (8.7) 21 (91.3)	 15 (7.2) 192 (92.8)	0.681
Lower extremity fracture Yes No	 3 (13) 20 (87)	 7 (3.4) 200 (96.6)	0.066
ISS group <16 ≥16	 18 (78.3) 5 (21.7)	 181 (87.4) 26 (12.6)	0.210
ISS $\bar{x} \pm$ SD	 15.5±20.0	 10.0±13.7	0.206
Length of hospitalization, (days) $\bar{x} \pm$ SD	 8.3±10.6	 5.2±12.2	0.243
Mortality Yes No	 3 (13) 20 (87)	 10 (4.8) 197 (95.2)	0.128

n: Number of patients, $\bar{x}$: Mean, SD: Standard deviation, ISS: Injury Severity score, categorical data were expressed as number of patients (percentage). Significant p-values as a result of statistical analysis were olded

Table 3. Low ISS and high ISS comparison			
	Low ISS	**High ISS**	**p**
	n=199	**n=31**	
Gender Male Female	 126 (63.3) 73 (36.7)	 19 (61.3) 12 (38.7)	0.828
Age, (year) $\bar{x} \pm SD$	 47±16.5	 52.1±20.4	0.277
Troponin, (ng/L) $\bar{x} \pm SD$	 0.01±0.05	 0.08±0.11	0.179
Trauma Vehicular traffic accident Non-vehicle traffic accident Assault Animal kick	 170 (85.4) 17 (8.5) 1 (0.5) 2 (1)	 19 (61.3) 7 (22.6) 0 (0) 2 (6.5)	0.007
Fracture localization Manubrium Sternal body Xiphoid process Mixed	 71 (35.7) 113 (56.8) 6 (3) 9 (4.5)	 13 (41.9) 16 (51.6) 0 (0) 2 (6.5)	0.713
Degree of fracture separation Non-displaced Displaced	 113 (56.8) 86 (43.2)	 15 (48.4) 16 (51.6)	0.381
Retrosternal hematoma Yes No	 18 (9) 181 (91)	 5 (16.1) 26 (83.9)	0.210
Isolated sternal fracture Yes No	 49 (24.6) 150 (75.4)	 0 (0) 31 (100)	0.002
Seat belt use Yes No	 61 (35.9) 109 (64.1)	 3 (15.8) 16 (84.2)	0.079
Rib fracture Yes No	 114 (57.3) 85 (42.7)	 26 (83.9) 5 (16.1)	0.005
Hemothorax Yes No	 14 (7) 185 (93)	 10 (32.3) 21 (67.7)	<0.001
Pulmonary contusion Yes No	 39 (19.6) 160 (80.4)	 11 (35.5) 20 (64.5)	0.046
Pneumothorax Yes No	 29 (14.6) 170 (85.4)	 13 (41.9) 18 (58.1)	<0.001
Pneumomediastinum Yes No	 2 (1) 197 (99)	 0 (0) 31 (100)	>0.999
Intracranial injury Yes No	 9 (4.5) 190 (95.5)	 13 (41.9) 18 (58.1)	<0.001
Maxillofacial trauma Yes No	 9 (4) 191 (96)	 7 (22.6) 24 (77.4)	0.001
Clavicle fracture Yes No	 11 (5.5) 188 (94.5)	 4 (12.9) 27 (87.1)	0.126

Table 3. Continued			
	Low ISS	**High ISS**	**p**
	n=199	**n=31**	
Scapular fracture Yes No	 9 (4.5) 190 (95.5)	 6 (19.4) 25 (80.6)	0.008
Vertebral fracture Yes No	 56 (28.1) 143 (71.9)	 15 (48.4) 16 (51.6)	0.023
Upper extremity fracture Yes No	 6 (3) 193 (97)	 4 (12.9) 27 (87.1)	0.032
Pelvic fracture Yes No	 13 (6.5) 186 (93.5)	 4 (12.9) 27 (87.1)	0.258
Intra-abdominal organ injury Yes No	 11 (5.5) 188 (94.5)	 6 (19.4) 25 (80.6)	0.016
Lower extremity fracture Yes No	 4 (2) 195 (98)	 6 (19.4) 25 (80.6)	0.001
Length of hospitalization, (days) $\bar{x} \pm$ SD	 4.68±11.0	 11.0±16.6	0.048
Mortality Yes No	 0 (0) 199 (100)	 13 (41.9) 18 (58.1)	<0.001

n: Number of patients, $\bar{x}$: Mean, SD: Standard deviation, ISS: Injury Severity score, categorical data were expressed as number of patients (percentage). Significant p-values as a result of statistical analysis were bolded

Table 4. Consequences of seat belt use			
	Seat belt Available	**Seat belt Not available**	**p**
	n=64	**n=125**	
Gender Male Female	 27 (42.2) 37 (57.8)	 52 (41.6) 73 (58.4)	0.938
Age, (year) $\bar{x} \pm$ SD	 50.4±14.3	 46.9±17.7	0.142
Fracture localization Manubrium Sternal body Xiphoid process Mixed	 15 (23.4) 44 (68.8) 3 (4.7) 2 (3.1)	 54 (43.2) 62 (49.6) 1 (0.8) 8 (6.4)	0.008
Degree of fracture separation Non-displaced Displaced	 32 (50) 32 (50)	 70 (54) 55 (44)	0.434
Retrosternal hematoma Yes No	 7 (10.9) 57 (89.1)	 12 (9.6) 113 (90.4)	0.772
Troponin, (ng/L) $\bar{x} \pm$ SD	 0.01±0.01	 0.02±0.05	0.405
Isolated sternal fracture Yes No	 19 (29.7) 45 (70.3)	 26 (20.8) 99 (79.2)	0.175
Rib fracture Yes No	 39 (60.9) 25 (39.1)	 74 (59.2) 51 (40.8)	0.818

Table 4. Continued			
	Seat belt available	**Seat belt not available**	**p**
	n=64	**n=125**	
Hemothorax Yes No	 4 (6.3) 60 (93.8)	 13 (10.4) 112 (89.5)	0.345
Pulmonary contusion Yes No	 10 (15.6) 54 (84.4)	 33 (26.4) 92 (73.6)	0.094
Pneumothorax Yes No	 6 (9.4) 58 (90.6)	 25 (20) 100 (80)	0.062
Pneumomediastinum Yes No	 0 (0) 64 (100)	 2 (1.6) 123 (98.4)	0.550
Intracranial injury Yes No	 1 (1.6) 63 (98.4)	 12 (9.6) 113 (90.4)	0.064
Maxillofacial trauma Yes No	 4 (6.3) 60 (93.8)	 9 (7.2) 116 (92.8)	>0.999
Clavicle fracture Yes No	 3 (4.7) 61 (95.3)	 4 (3.2) 121 (96.8)	0.690
Scapular fracture Yes No	 1 (1.6) 63 (98.4)	 8 (6.4) 117 (93.6)	0.277
Upper extremity fracture Yes No	 2 (3.1) 62 (96.9)	 8 (6.4) 117 (93.6)	0.499
Vertebral fracture Yes No	 11 (17.2) 53 (82.8)	 40 (32) 85 (68)	0.030
Pelvic fracture Yes No	 2 (3.1) 62 (96.9)	 8 (6.4) 117 (93.6)	0.499
Intra-abdominal organ injury Yes No	 2 (3.1) 62 (96.9)	 12 (9.6) 113 (90.4)	0.145
Lower extremity fracture Yes No	 3 (4.7) 61 (95.3)	 5 (4) 120 (96)	>0.999
ISS $\bar{x} \pm$ SD	 5.3±5.1	 10.6±14.4	0.005
ISS group <16 ≥16	 61 (95.3) 3 (4.7)	 109 (87.2) 16 (12.8)	0.079
Length of hospitalization, (days) $\bar{x} \pm$ SD	 2.8±2.7	 4.9±6.0	0.001
Mortality Yes No	 0 (0) 64 (100)	 119 (95.2) 6 (4.8)	0.098

n: Number of patients, $\bar{x}$: Mean, SD: Standard deviation, ISS: Injury Severity score, categorical data were expressed as number of patients (percentage). Significant p-values as a result of statistical analysis were bolded

Discussion

This study demonstrated the incidence of sternal fractures in individuals with thoracic trauma at 10.7%. Literature indicates that the frequency of sternal fractures in patients with thoracic trauma varies between 3% and 8% (6,10,22). Recent developments in imaging technologies may account for the elevated rate observed in our study. The hospital's proximity to a busy road that links major centers may increase the incidence of accidents, thereby resulting in a higher observed rate.

The predominant cause of sternal fractures was traffic accidents (82.2%). The enhanced capability of cars to attain greater velocities is a critical contributor to the escalating incidence of traffic accidents (15). Although seatbelt utilization decreases mortality in vehicular collisions, it has been linked to a heightened occurrence of sternal fractures. Literature indicates that sternal fractures occur 0.7% to 4% more frequently in belted passengers (15). Sternal fractures are frequently termed "seatbelt trauma" (5-7,10,15). In our study, 33.9% of patients with sternal fractures resulting from traffic accidents were utilizing seatbelts. The absence of mortality in these patients, their shorter hospitalization, and a lower average ISS further support findings reported in the literature.

Our analysis predominantly identified non-displaced corpus fractures. The site and classification of sternal fractures are crucial for diagnosing thoracic organ damage (14,15,23). Research indicates that the prevalence of cardiac injury in individuals with sternal fractures varies between 1% and 21% (10,24). The resultant mortality is predominantly linked to corpus fractures (22-25). Our results were largely consistent with the existing literature, albeit minor statistical differences, which we attribute to the limited sample size.

The most commonly detected thoracic trauma associated with sternal fractures was rib fractures (n=140, 54.3%),while the predominant extra-thoracic injury was vertebral fractures (n=71, 36.9%). Literature frequently indicates an association between sternal fractures and rib fractures (14). Nevertheless, no prior research has specifically examined the association between sternal fractures and extra-thoracic injuries. Our data were consistent–with the literature, and we contend that the relationship with extra-thoracic injuries may fluctuate depending on the mechanism of trauma.

The literature underscores the significance of electrocardiography and cardiac enzyme evaluation in diagnosing cardiac injury. We advocate performing echocardiogram upon identification of abnormal findings. Several studies emphasize the necessity of hospitalization for patients exhibiting abnormal electrocardiographic findings associated with sternal fractures, whereas others suggest that hospitalization may be unnecessary in patients with only mildly elevated cardiac enzyme levels (23,26). In our study, all patients underwent electrocardiographic evaluation; however, no abnormalities were detected. Consequently, additional investigations were not pursued based solely on electrocardiographic findings. Troponin levels were assessed in patients suspected of cardiac injury during thoracic CT evaluation, irrespective of electrocardiography results, and echocardiography was performed in those with elevated levels. Although echocardiography revealed no evidence of cardiac injury, a positive association was observed between troponin levels and ISS. Despite differing perspectives in the literature, we advocate hospitalization and close monitoring of patients with elevated troponin levels in the presence of sternal fractures.

The management of sternal fractures depends on the degree of fracture displacement and the presence of associated complications; nonetheless, conservative treatment remains the primary approach in most cases (27). In our investigation, a single patient developed hemothorax secondary to a retrosternal hematoma, and was managed with VATS and hematoma evacuation. This patient had no associated cardiac or vascular injuries. The origin of the retrosternal hematoma was attributed to be the mediastinal fascias and the sternal structures. Literature indicates that retrosternal hematoma occurs in approximately 25% of patients with sternal fractures, a prevalence consistent with previous reports (28). In our study, however, this rate was lower, and the presence of retrosternal hematoma was not associated with the length of hospitalization, mortality, or trauma severity scores. Consistent with the literature, retrosternal hematoma was not associated with cardiac injury (29,30). However, its significant association with pulmonary contusion and hemothorax in our study suggests that clinicians should remain vigilant for potential pulmonary complications in patients with sternal fractures.

Sternal fractures represent a substantial source of morbidity and mortality due to their frequent association with concomitant organ injuries. Literature reports suggest that mortality rates in these patients are approximately 10% (14,15). In our study, however, the mortality rates was lower than that reported previously. Multiple factors may influence mortality, with early initiation of appropriate treatment and the quality of healthcare delivery being critical determinants. We believe that these factors contributed to the favorable outcomes observed in our cohort. Nevertheless, further studies are needed to validate these findings.

Study Limitations

The limitations of our study include its retrospective nature, single-center design, and relatively small sample size. Reliance on patient self-report to obtain information regarding seatbelt use may be considered a limitation of this study. In addition, the absence of a radiologist in the evaluation of radiological imaging and image interpretation being conducted by a single experienced thoracic surgeon may also be regarded as study limitations.

Conclusion

Seatbelt use in traffic accidents is life-saving; however, it can lead to rare injuries such as sternal fractures. While conservative treatment often successfully manages sternal fractures, signs such as the presence of a retrosternal hematoma may suggest associated pulmonary injury rather than cardiac involvement. The use of appropriate diagnostic modalities and close in-hospital monitoring of these patients can help prevent mortality.

Ethics

Ethics Committee Approval: Ethical approval was obtained from Ethics Committee Approval for Non-Interventional Research from Kırıkkale University on (decision number: 2021.07.02, date: 08.07.2021). The study was conducted in accordance with the Declaration of Helsinki.

Informed Consent: The study is notable for its retrospective nature, single-center design, and relatively small sample size.

Footnotes

Authorship Contributions

Surgical and Medical Practices: S.K., E.Z., N.G., Concept: S.K., N.G., Design: S.K., N.G., Data Collection or Processing: S.K., K.B.Ç., E.Z., N.G., Analysis or Interpretation: S.K., K.B.Ç., E.Z., N.G., Literature Search: S.K., K.B.Ç., E.Z., N.G., Writing: S.K., K.B.Ç., E.Z., N.G.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. LoCicero J, Mattox KL. Epidemiology of chest trauma. Surg Clin North Am 1989;69:15-9.
2. Demirhan R, Onan B, Öz K, Halezeroglu S. Comprehensive analysis of 4205 patients with chest trauma: a 10-year experience. Interact Cardiovasc Thorac Surg. 2009;9:450-3.
3. Mirka H, Ferda J, Baxa J. Multidetector computed tomography of chest trauma: indications, technique and interpreta-tion. Insights Imaging. 2012;3:433-49.
4. Oikonomou A, Prassopoulos P. CT imaging of blunt chest trauma. Insights Imaging. 2011;2:281-95.
5. Knobloch K, Wagner S, Haasper C, Probst C, Krettek C, Otte D, et al. Sternal fractures occur most often in old car sto seat-belted drivers without any airbag often with concomitant spinal injuries: clinical findings and technical collision variables among 42,055 crash victims. Ann Thorac Surg. 2006;82:444-50.
6. Athanassiadi K, Gerazounis M, Moustardas M, Metaxas E. Sternal fractures: retrospective analysis of 100 cases. World J. Surg. 2002;26:1243-6.
7. Sadaba JR, Oswal D, Munsch CM. Management of isolated sternal fractures: determining the risk of blunt cardiac injury. Ann R Coll Surg Engl. 2000;82:162-6.
8. Brookes JG, Dunn RJ, Rogers IR. Sternal fractures: a retrospective analysis of 272 cases. J Trauma. 1993;35:46-54.
9. Çetin M, Solak N, Özdemir Çiflik B, Türk İ, Ermancik SS, Aydoğdu K. Thoracic trauma in the geriatric population and possible preventive measures: a retrospective analysis of 261 cases. Turk J Med Sci. 2024;54:1013-20.
10. Çelik B, Şahin E, Nadir A, Kaptanoglu M. Sternum fractures and effects of associated injuries. Thorac Cardiovasc Surg. 2009;57:468-71.
11. Gavelli G, Canini R, Bertaccini P, Battista G, Bnà C, Fattori R. Traumatic injuries: imaging of thoracic injuries. Eur Radiol. 2002;12:1273-94.
12. Recinos G, Inaba K, Dubose J, Barmparas G, Teixeira PG, Talving P, et al. Epidemiology of sternal fractures. Am Surg. 2009;75:401-4.
13. Lee NH, Kim SH, Kim JH, Kim HH, Lee SB, Park CI, et al. A decade of treating traumatic sternal fractures in a single-center experience in Korea: a retrospective cohort study. J Trauma Inj. 2023;36:362-8.
14. Aamir J, Alade B, Caldwell R, Chapman J, Shah S, Karthikappallil D, et al. Sternal fractures and thoracic injury: an analysis of 288 sternal fractures attending a major trauma centre. Eur J Orthop Surg Traumatol. 2023;33:2619-624.
15. Şimşek S, Özmen CA, Onat S. Morbidity and mortality associated with fracture of the sternum due to blunt trauma, by fracture type and location. Radiol Bras. 2022;55:167-172.
16. Chapman BC, Overbey DM, Tesfalidet F, Schramm K, Stovall RT, French A, et al. Clinical utility of chest computed tomography in patients with rib fractures CT chest and rib fractures. Arch Trauma Res. 2016;5:e37070.
17. Flagel BT, Luchette FA, Reed RL, Esposito TJ, Davis KA, Santaniello JM, et al. Half-a-dozen ribs: the breakpoint for mortality. Surgery. 2005;138:717-23; discussion 723-5.
18. Son YN, Kim JI, Lee HN, Shin SY. Retrosternal hematoma in sternal fracture for prediction of concomitant injury on chest CT. Acta Radiol. 2021;62:1610-7.
19. Van Diepen MR, Wijffels MME, Verhofstad MHJ, Van Lieshout EMM. Classification methods of pulmonary contusion based on chest CT and the association with in-hospital outcomes: a systematic review of literature. Eur J Trauma Emerg Surg. 2024;50:2727-40.
20. Huang JF, Hsu CP, Fu CY, Ou Yang CH, Cheng CT, Liao CH, et al. Is massive hemothorax still an absolute indication for operation in blunt trauma? Injury. 2021;52:225-30.
21. Grewal J, Gillaspie EA. Pneumomediastinum. Thorac Surg Clin. 2024;34:309-19.
22. Turhan K, Çakan A, Özdil A, Çağırıcı U. Traumatic sternal fractures: diagnosis and management. ETD. 2010; 49:107-11.
23. Esme H, Büyükterzi Z. Risk factors of blunt cardiac injury and routine use of echocardiography in sternum fractures. JSurgArts. 2020;13:53-7.
24. Potkin RT, Werner JA, Trobaugh GB, Chestnut CH, Carrico CJ, Hallstrom A, et al. Evaluation of non-invasive tests of cardiac damage in suspected cardiac contusion. Circulation. 1982;66:627-31.

25. Von Garrel T, Ince A, Junge A, Chestnut CH, Carrico CJ, Hallstrom A, et al. The sternal fracture: radiographic analysis of 200 fractures with special reference to concomitant injuries. J Trauma 2004;57:837-44.
26. Al-Khouja F, Grigorian A, Emigh B, Schellenberg M, Diaz G, Duncan TK, et al. 24-hour telemetry monitoring may not be necessary for patients with an ısolated sternal fracture and minor ECG abnormalities or troponin elevation: a southern california multicenter study. Am Surg. 2025;91:126-32.
27. Xu S, Zhu J, Yu Q, Peng L, Tao Y, Qi S, et al. Surgical treatment of sternum comminuted fracture with memory alloy embracing fixator. J Thorac Dis. 2021;13:2194-202.
28. Fokin AA, Wycech Knight J, Yoshinaga K, Abid AT, Grady R, Alayon AL, et al. Blunt cardiac injury in patients with sternal fractures. Cureus. 2022;14:e22841.
29. Heidelberg L, Uhlich R, Bosarge P, Kerby J, Hu P. The depth of sternal fracture displacement is not associated with blunt cardiac injury. J Surg Res. 2019;235:322-8.
30. Rashid MA, Ortenwall P, Wikström T. Cardiovascular injuries associated with sternal fractures. Eur J Surg. 2001;167:243-8.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.46548
Eurasian J Emerg Med. 2026;25: 222-7

The Role of the Lactate Dehydrogenase-to-albumin Ratio in Differentiating Acute Renal Infarction from Ureterolithiasis

Nazım Abdülkadir Kankılıç[1], Ekrem Taha Sert[2], Kamil Kokulu[2]

[1]Aksaray University Faculty of Medicine, Department of Urology, Aksaray, Türkiye
[2]Aksaray University Faculty of Medicine, Department of Emergency Medicine, Aksaray, Türkiye

Abstract

Aim: Acute renal infarction (ARI) is an uncommon but clinically important vascular emergency that is frequently misdiagnosed as acute ureterolithiasis due to overlapping symptoms. This study aimed to evaluate the diagnostic value of the lactate dehydrogenase-to-albumin (LDH/albumin) ratio in patients with ARI.

Materials and Methods: This study included adult patients diagnosed with ARI between January 2018 and June 2025. ARI was confirmed using contrast-enhanced computed tomography. A control group of patients with acute ureterolithiasis was selected using 2:1 propensity score matching on age and sex. Demographic variables, clinical features, and laboratory parameters were compared. Univariate and multivariate logistic regression analyses were performed to identify independent predictors of ARI. Diagnostic performance was assessed using ROC curve analysis.

Results: A total of 162 patients were included, comprising 54 with ARI and 108 with ureterolithiasis. Patients with ARI showed significantly higher LDH levels and LDH/albumin ratios, and lower albumin levels compared with controls. The LDH/albumin ratio demonstrated the strongest association with ARI [odds ratio=3.62; 95% confidence interval (CI): 1.96-6.58]. ROC analysis showed that the LDH/albumin ratio [area under the curve (AUC)=0.921; 95% CI: 0.881-0.963) had superior diagnostic performance compared with LDH (AUC=0.873) and albumin (AUC=0.780). A cut-off value >13.1 yielded a sensitivity of 85.2% and a specificity of 83.3% for identifying ARI.

Conclusion: The LDH/albumin ratio demonstrates superior diagnostic performance compared with LDH or albumin alone because it reflects both tissue ischemia and inflammatory processes. Therefore, it may serve as a valuable and readily accessible biomarker for the early identification of ARI.

Keywords: Acute renal infarction, lactate dehydrogenase, albumin, diagnostic value

Introduction

Acute renal infarction (ARI) is a rare but clinically significant vascular condition resulting from complete or partial occlusion of the renal artery, typically due to embolic or thrombotic events. Delayed diagnosis carries a significant risk of permanent renal damage (1). The low diagnostic rate of ARI in the emergency department (ED) is primarily explained by its non-specific clinical presentation, which frequently leads to misdiagnosis as urolithiasis. These diagnostic challenges hinder accurate estimation of the true incidence of ARI. The most common underlying etiologies include atrial fibrillation, heart failure, and valvular heart diseases (2). ARI typically presents with acute abdominal or flank pain, often accompanied by characteristic laboratory findings, including elevated serum lactate dehydrogenase (LDH), microscopic hematuria, and elevated creatinine levels (3).

The LDH, a non-specific marker of cellular injury, may rise in a variety of clinical conditions, including myocardial infarction, hemolysis, and acute ischemic renal tubular necrosis (4,5). Therefore, LDH measurement alone is not specific enough to reliably establish the diagnosis of ARI. To improve diagnostic accuracy, LDH should be evaluated in conjunction with additional biochemical parameters. Recently, the LDH/albumin

Corresponding Author: Nazım Abdülkadir Kankılıç MD, Aksaray University Faculty of Medicine, Department of Urology, Aksaray, Türkiye
E-mail: nakankilicdr@hotmail.com **ORCID ID:** orcid.org/0000-0002-3747-3798

Cite this article as: Kankılıç NA, Sert ET, Kokulu K. The role of the lactate dehydrogenase-to-albumin ratio in differentiating acute renal infarction from ureterolithiasis. Eurasian J Emerg Med. 2026;25: 222-7.

Received: 22.01.2026
Accepted: 03.03.2026
Published: 13.03.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

ratio has been investigated as a prognostic indicator in various conditions characterized by inflammation and tissue injury. In particular, an elevated LDH/albumin ratio has been identified as an independent predictor of mortality in sepsis-associated acute kidney injury (AKI) (6) and has also been reported as a valuable marker for prognostic assessment in critically ill patients with AKI (7). Current evidence suggests that the LDH/albumin ratio may play a role in the diagnosis of ARI, serving as an indicator of cellular injury and inflammation. However, the diagnostic utility of the LDH/albumin ratio for ARI remains poorly defined in the existing literature. Therefore, this study aimed to evaluate the diagnostic value of the LDH/albumin ratio in patients with ARI.

Materials and Methods

Study Design and Patient Selection

This single-center, retrospective, case-control study used data from patients diagnosed with ARI in the ED between January 2018 and June 2025. Inclusion criteria were defined as adult patients (≥18 years) with ARI confirmed by contrast-enhanced abdominal computed tomography (CT). Data extraction was performed through a retrospective review of electronic medical records. The study protocol was approved by the Aksaray University Health Sciences Scientific Research Ethics Committee (decision number: 2025/006, date: 16.01.2025), and the study was conducted in accordance with the principles of the Declaration of Helsinki.

Patients were excluded if they met any of the following criteria: (1) age <18 years; (2) ARI not confirmed by contrast-enhanced CT; (3) insufficient clinical, laboratory, or imaging data; (4) absence of abdominal or flank pain; (5) history of acute or chronic liver disease; (6) presence of sepsis or systemic inflammatory disease; (7) history of malignancy; (8) history of chronic kidney disease or prior renal transplantation; or (9) clinical conditions known to significantly affect LDH levels, such as major trauma, rhabdomyolysis, or myocardial infarction.

Data Collection

Patient demographic characteristics (age, sex), presenting vital signs, chief complaint, comorbidities (e.g., atrial fibrillation, hypertension, diabetes mellitus), and the anatomical location of ARI (right or left kidney) were obtained from electronic medical records. All laboratory parameters, including LDH and albumin levels, were obtained from venous blood samples collected at the time of presentation to the ED. Contrast-enhanced abdominal CT images were obtained using a 128-slice GE Revolution EVO (USA) CT scanner. The ARI group was defined based on contrast-enhanced abdominal CT findings consistent with ARI. The final diagnosis of ARI was established by consensus between an emergency medicine specialist and a urologist, based on a review of the radiologists' abdominal CT reports and the corresponding images. All demographic, laboratory, and imaging data were systematically recorded using a standardized, study-specific data collection form.

The control group consisted of patients with acute ureterolithiasis who presented to the ED with similar symptoms (flank or abdominal pain) during the same period; ARI was excluded by contrast-enhanced CT performed as part of their diagnostic evaluation. To increase the statistical power of the study, the control group was matched to the case group at a 2:1 ratio on age and sex using propensity score matching. This approach reduced the influence of potential confounding variables, thereby enabling a more reliable assessment of the independent diagnostic utility of the LDH/albumin ratio in ARI. The matching process was performed using the nearest-neighbor algorithm with a caliper width of 0.2. The effectiveness of the propensity score matching was confirmed by standardized mean difference values <0.1, indicating a well-balanced distribution between the matched groups. This approach substantially reduced bias and enhanced the reliability of the evaluation of the diagnostic performance of the laboratory indicators.

Statistical Analysis

Statistical analyses were performed using SPSS version 22.0 (SPSS Inc., Chicago, IL, USA). Categorical variables were presented as frequencies and percentages, and comparisons were conducted using the chi-square test or Fisher's exact test. The distribution of continuous variables was assessed using the Kolmogorov-Smirnov test. Normally distributed continuous data were compared using the Student's t-test and presented as mean ± standard deviation, whereas non-normally distributed data were compared using the Mann-Whitney U test and expressed as median (25^{th}-75^{th} percentile). To investigate the independent association between the LDH/albumin ratio and ARI, a univariate logistic regression analysis was first performed; variables with a p-value <0.05 were then included in a multivariate logistic regression model. Odds ratios (ORs) and 95% confidence intervals (CIs) were reported. The diagnostic performance of the LDH/albumin ratio for ARI was evaluated using ROC curve analysis. The area under the curve (AUC) and its 95% CI were calculated, and the optimal cut-off value was determined using the Youden index. ROC curve AUCs were compared using the DeLong test to assess differences in diagnostic performance between the LDH/albumin ratio and individual biomarkers. Sensitivity, specificity, and positive and negative predictive values were also calculated. For all analyses, a p-value <0.05 was considered statistically significant.

Results

A total of 162 patients who met the eligibility criteria were included in the study. Of these, 108 patients (66.7%) were assigned to the control group (acute ureterolithiasis), and 54 patients (33.3%) were diagnosed with ARI. The median age was 50 (46-55) years in the control group and 53 (47-58) years in the ARI group. Male patients constituted 53.7% of the control group and 55.6% of the ARI group. Flank pain was significantly more common in the control group (80.6% vs. 51.9%; p<0.001), whereas abdominal pain was observed more frequently in the ARI group (74.1% vs. 47.2%; p=0.001). The prevalence of atrial fibrillation was markedly higher in the ARI group compared with controls (27.8% and 9.3%, respectively; p=0.002). Patients with ARI also exhibited significantly elevated white blood cell (WBC) counts (11.7±2.7 and 10.5±1.9, respectively; p=0.001), LDH levels (680±148 and 468±99 U/L, respectively; p<0.001), and LDH/albumin ratio (16.5±3.4 and 10.5±2.3, respectively; p<0.001) (Figure 1). In contrast, albumin levels were lower in the ARI group (41.2±2.9 and 44.6±3.6 g/L, respectively; p<0.001), while hematuria was more frequently observed among control patients (64.8% and 42.6%, respectively; p=0.007) (Table 1).

Abdominal pain OR=2.66; CI=1.55-4.56; p=0.001, atrial fibrillation (OR=3.73; CI=1.62-8.01; p=0.002), and WBC count (OR=1.19; CI=1.05-1.38; p=0.003) were all significantly associated with a higher likelihood of ARI. Among biochemical markers, LDH demonstrated a strong association

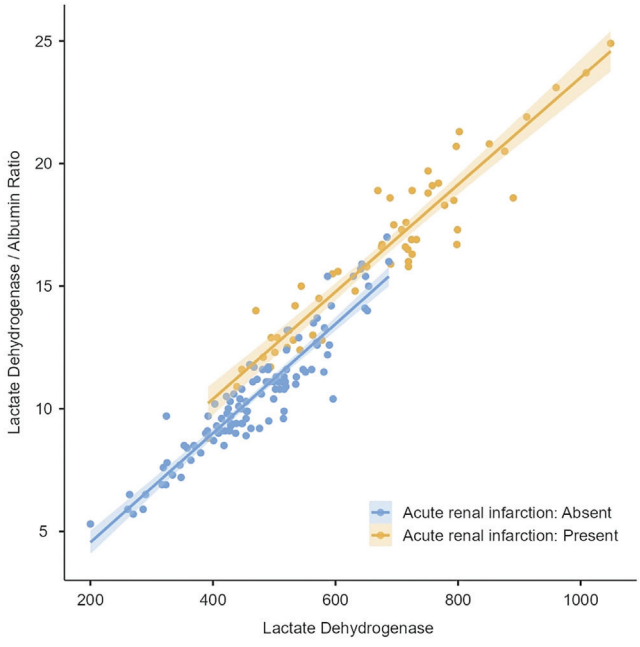


Figure 1. Distribution of patients according to LDH Levels and LDH/albumin ratios by acute renal infarction status

LDH: Lactate dehydrogenase

Table 1. Comparison of patient characteristics between the acute renal infarction group and the acute ureterolithiasis (Control) group

Variables	Control group (n=108)	ARI group (n=54)	p-value
Age, years	50 (46-55)	53 (47−58)	0.106
Male sex	58 (53.7%)	30 (55.6%)	0.823
Vital sign			
SBP, mmHg	132.7±12.1	133.5±12.2	0.704
DBP, mmHg	81.9±6.4	82.6±7.1	0.462
Heart rate, bpm	84 (80-92)	85 (81-94)	0.089
Body temperature, °C	36.8±0.3	36.9±0.4	0.323
Presenting symptoms			
Abdominal pain	51 (47.2%)	40 (74.1%)	0.001
Flank pain	87 (80.6%)	28 (51.9%)	<0.001
Nausea/vomiting	43 (39.8%)	26 (48.1%)	0.312
Fever	6 (5.6%)	5 (9.3%)	0.377
Other	3 (2.8%)	2 (3.7%)	0.748
Past medical history			
Hypertension	31 (28.7%)	17 (31.5%)	0.715
Diabetes mellitus	17 (15.7%)	11 (20.4%)	0.463
Cardiovascular disease	8 (7.4%)	6 (11.1%)	0.429
Atrial fibrillation	10 (9.3%)	15 (27.8%)	0.002
Dyslipidemia	6 (5.6%)	8 (14.8%)	0.072
Ureteral stone	41 (38.0%)	10 (18.5%)	0.012
Side of ARI on CT			
Right	50 (46.3%)	29 (53.7%)	0.374
Left	58 (53.7%)	25 (46.3%)	
Laboratory data (reference range)			
WBC, (4-10 103/mL)	10.5±1.9	11.7±2.7	0.001
Hemoglobin, (11-16 g/dL)	13.0±1.1	13.1±1.2	0.626
Platelet count, (100-400 10^9/L)	246±44	241±31	0.578
Urea, (12-43 mg/dL)	25.5±7.3	27.3±7.4	0.104
Creatinine, (0.67-1.17 mg/dL)	0.89±0.30	0.93±0.29	0.215
AST, (0-50 U/L)	36.4±10.5	38.8±9.1	0.118
ALT, (0-50 U/L)	34.4±9.2	35.1±8.4	0.433
Albumin, (35-52 g/L)	44.6±3.6	41.2±2.9	<0.001
LDH, (0-248U/L)	468±99	680±148	<0.001
CRP, (0-5 mg/L)	3.8±2.9	3.2±2.5	0.456
LDH/albumin ratio	10.5±2.3	16.5±3.4	<0.001
Urinalysis			
Proteinuria	17 (15.7%)	8 (14.8%)	0.878
Hematuria	70 (64.8%)	23 (42.6%)	0.007

Data are expressed as mean ± standard deviation, median (25^{th}-75^{th} quartile) and percentiles or n (%). ARI: Acute renal infarction, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, WBC: White blood cell count, AST: Aspartate aminotransferase, ALT: Alanine aminotransferase, CRP: C-reactive protein, LDH: Lactate dehydrogenase, CT: Computed tomography

(OR=2.91; CI=1.80-4.72; p<0.001), while the LDH/albumin ratio emerged as the most powerful predictor (OR=4.85; CI=3.00-7.72; p<0.001). Conversely, flank pain (OR=0.27; CI=0.16-0.44; p<0.001), a history of ureteral stones (OR=0.38; CI=0.18-0.74; p=0.014), lower albumin levels (OR=0.79; CI=0.66-0.94; p<0.001), and hematuria (OR=0.43; CI=0.25-0.70; p=0.006) were significantly associated with reduced odds of ARI (Table 2). Independent predictors of ARI were evaluated using multivariate logistic regression analysis (Table 3). To avoid multicollinearity, LDH and albumin were excluded from the multivariate model, and the LDH/albumin ratio remained the strongest independent predictor of ARI (OR=3.62; CI=1.96-6.58; p<0.001). Atrial fibrillation was the only clinical comorbidity that retained an independent and statistically significant association with ARI (OR=2.46; CI=1.12–5.05; p=0.021). The diagnostic performance of LDH, albumin, and the LDH/albumin ratio in identifying ARI was evaluated using ROC curve analysis (Table 4). The LDH/albumin ratio demonstrated the highest discriminative ability, with an AUC of 0.921 (CI=0.881-0.963), followed by LDH with an AUC of 0.873 (CI=0.820–0.934), and albumin with an AUC of 0.780 (CI=0.711-0.853). A cut-off value of >13.1 for the LDH/albumin ratio yielded a sensitivity of 85.2% and a specificity of 83.3%, indicating strong diagnostic accuracy for distinguishing ARI from ureterolithiasis.

Table 2. Unadjusted associations between clinical and laboratory variables and the presence of acute renal infarction

Variables	OR (95% CI)	p-value
Abdominal pain	2.66 (1.55-4.56)	0.001
Flank pain	0.27 (0.16-0.44)	<0.001
Atrial fibrillation	3.73 (1.62-8.01)	0.002
Ureteral stone	0.38 (0.18-0.74)	0.014
WBC	1.19 (1.05-1.38)	0.003
Albumin	0.79 (0.66-0.94)	<0.001
LDH	2.91 (1.80-4.72)	<0.001
LDH/album ratio	4.85 (3.00-7.72)	<0.001
Hematuria	0.43 (0.25–0.70)	0.006

OR: Odds ratio, CI: Confidence interval, WBC: White blood cell count, LDH: Lactate dehydrogenase

Table 3. Independent predictors of acute renal infarction in multivariate analysis

Variables	OR (95% CI)	p-value
Abdominal pain	1.28 (0.63-2.11)	0.218
Flank pain	0.74 (0.40-1.26)	0.193
Atrial fibrillation	2.46 (1.12-5.05)	0.021
Ureteral stone	0.91 (0.48-1.68)	0.613
WBC	1.07 (0.95-1.22)	0.264
LDH/albumin ratio	3.62 (1.96-6.58)	<0.001
Hematuria	0.82 (0.45-1.50)	0.493

OR: Odds ratio, CI: Confidence interval, WBC: White blood cell count, LDH: Lactate dehydrogenase

Discussion

ARI is a clinically uncommon condition, and current data indicate that an accurate diagnosis is established in only about 40% of patients at the time of presentation (8,9). This low diagnostic rate stems from the non-specific clinical presentation, with symptoms such as sudden-onset flank or abdominal pain, nausea, vomiting, and hematuria (10-12). These symptoms are frequently mistaken for more common clinical conditions, such as nephrolithiasis, ureterolithiasis, cholelithiasis, pyelonephritis, or gastrointestinal infections, and may therefore lead to delays in diagnosis. In this context, clinicians' ability to recognize clinical and laboratory findings suggestive of ARI may play a critical role in facilitating early diagnosis and reducing the risk of irreversible renal injury. Thus, the identification of novel biomarkers that may contribute to the diagnosis of ARI is of considerable clinical importance. Accordingly, this study investigated the diagnostic utility of the LDH/albumin ratio in patients with ARI. We found the LDH/albumin ratio to be the strongest independent predictor of ARI, and it demonstrated the highest diagnostic performance, with an AUC of 0.921 in differentiating ARI from acute ureterolithiasis.

Ischemia, cellular hypoxia, oxidative stress, and the inflammatory response during ARI substantially increase the release of LDH into the plasma. Oxygen deficiency within renal tissue triggers the LDH-mediated conversion of pyruvate to lactate, shifting energy production toward anaerobic metabolism and thereby increasing cellular metabolic stress. Elevated LDH may further promote oxidative stress, while the resulting lactic acid can exacerbate inflammation by influencing the production of nitric oxide and cytokines such as TNF-α and interleukin-6 (13,14). LDH, a marker of cellular necrosis, rises rapidly during renal infarction and may reach levels four to five times the normal level. Although it gradually declines after symptom onset, it typically remains elevated for approximately 15 days (5). However, because LDH levels vary depending on the size and duration of the infarction, establishing a reliable cut-off value based on LDH alone is challenging. Because LDH is present in multiple tissues, various forms of cellular injury can elevate serum LDH levels. This may lead to both false-positive and false-negative interpretations, thereby limiting its diagnostic utility (15). In the literature, elevated serum LDH levels have been reported in more than 70% of patients with ARI (1,16-18). Previous studies have demonstrated associations between ARI and both a history of atrial fibrillation and elevated LDH levels (16,17). Similarly, Zhang and Liu (19) reported that LDH levels

Table 4. Receiver operating characteristic curve analysis for the diagnostic performance of LDH, albumin, and the LDH/albumin ratio in identifying acute renal infarction

Variables	Albumin	LDH	LDH/albumin ratio
AUC (95% CI)	0.780 (0.711-0.853)	0.873 (0.820-0.934)	0.921 (0.881-0.963)
Cut-off value	<41 g/L	>585 U/L	>13.1
Sensitivity (%)	72.2	80.0	85.2
Specificity (%)	70.4	79.6	83.3
+LR	2.44	3.92	5.10
-LR	0.39	0.25	0.18
PPV (%)	57.0	67.5	73.0
NPV (%)	81.5	88.4	92.0

AUC: Area under the curve, CI: Confidence interval, LR: Likelihood ratio, PPV: Positive predictive value, NPV: Negative predictive value, LDH: Lactate dehydrogenase

were frequently elevated in ARI and suggested that this marker may help differentiate ARI from other causes of acute flank pain, including renal colic. Additionally, elevations in aspartate aminotransferase, alanine aminotransferase, and C-reactive protein (CRP) levels have also been reported to be associated with ARI (17,20). In a retrospective study comparing renal infarction and acute ureteral stone cases in the ED, Ahn and Lee (18) reported that advanced age (≥70 years), a history of atrial fibrillation, fever (≥37.5 °C), elevated LDH (≥500 IU/L), low chloride (≤103 mEq/L), low albumin (≤4.3 g/dL), elevated CRP (≥0.23 mg/dL), and the absence of urine erythrocytes were all significan tly associated with an increased probability of ARI. Similarly, in our study LDH levels were significantly higher in the ARI group.

Albumin plays a critical role in renal tissue due to its antioxidant and anti-inflammatory properties. It also helps maintain colloid osmotic pressure, thereby supporting fluid balance, intravascular volume, and tissue perfusion (21-24). Albumin, which interacts with oxidants through its free thiol group, may decrease in plasma levels during acute stress responses, leading to hypoalbuminemia (25-27). The reduction in albumin levels not only contributes to tissue injury by compromising renal endothelial stability but also diminishes the kidney's resistance to oxidative and inflammatory stress by limiting albumin's antioxidant and anti-inflammatory effects (13,28). Therefore, the simultaneous assessment of LDH and albumin may serve as a valuable indicator for the early diagnosis of ARI. The LDH/albumin ratio has recently been recognized as a novel diagnostic and prognostic biomarker across several clinical conditions. Fang et al. (6) reported that the LDH/albumin ratio is an independent risk factor for the development of AKI in patients with sepsis. Mutailifu et al. (13) demonstrated that a high LDH/albumin ratio is associated with in-hospital mortality in aortic dissection, while Chu et al. (29) showed that it has prognostic value for predicting adverse clinical outcomes in patients with ischemic stroke. In our study, we demonstrated that the LDH/albumin ratio was the strongest independent predictor of ARI and showed excellent diagnostic performance, with an AUC of 0.921 for distinguishing ARI from acute ureterolithiasis. These results suggest that including albumin alongside LDH may increase diagnostic accuracy and that albumin may serve as a valuable diagnostic biomarker for clinical decision-making in patients presenting with acute flank pain.

Study Limitations

This study has several limitations. First, it was conducted retrospectively at a single center, which may limit generalizability to broader patient populations. Second, only patients with available LDH and albumin measurements were included, and these biomarkers were assessed at a single time point; therefore, changes over time could not be evaluated. Third, the relatively small number of patients diagnosed with ARI reduced the statistical power. Fourth, the diagnosis of ARI relied exclusively on contrast-enhanced CT; consequently, patients unable to undergo contrast-enhanced CT were excluded, potentially introducing selection bias. Fifth, LDH and albumin levels are influenced by several external factors, including nutritional status, chronic illnesses, and nonspecific tissue injury, all of which may have introduced unmeasured confounding. Another limitation of this study is that serum LDH levels may be influenced by hemolysis; because of the retrospective design, hemolyzed blood samples could not always be reliably identified. Finally, using acute ureterolithiasis alone as the control group limits the generalizability of the LDH/albumin ratio's diagnostic performance to other causes of acute abdominal or flank pain.

Conclusion

Our findings indicate that the LDH/albumin ratio provides superior diagnostic accuracy compared with LDH or albumin alone, likely because it reflects both tissue ischemia and underlying inflammatory processes. Consequently, the LDH/albumin ratio may serve as a valuable and readily accessible biomarker for the early identification of ARI. Further studies are

required to validate its clinical utility and determine its role in routine diagnostic practice.

Ethics

Ethics Committee Approval: The study protocol was approved by the Aksaray University Health Sciences Scientific Research Ethics Committee (decision number: 2025/006, date: 16.01.2025), and the study was conducted in accordance with the principles of the Declaration of Helsinki.

Informed Consent: Informed consent was waived due to the retrospective study design.

Footnotes

Authorship Contributions

Surgical and Medical Practices: N.A.K., K.K., E.T.S., Concept: N.A.K., Design: E.T.S., Data Collection or Processing: N.A.K., Analysis or Interpretation: N.A.K., K.K., Literature Search: N.A.K., Writing: N.A.K.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Çağdaş Ş, Asfuroğlu A, Aykanat İC, Özercan AY, Köseoğlu B, Balcı M, et al. Acute renal infarction: a single-center experience. Eur J Ther. 2022;28:209-13.
2. Oh YK, Yang CW, Kim YL, Kang SW, Park CW, Kim YS, et al. Clinical characteristics and outcomes of renal infarction. Am J Kidney Dis. 2016;67:243-50.
3. Ongun S, Bozkurt O, Demir O, Cimen S, Aslan G. Midterm renal functions following acute renal infarction. Kaohsiung J Med Sci. 2015;31:529-33.
4. Muchtar E, Dispenzieri A, Lacy MQ, Buadi FK, Kapoor P, Hayman SR, et al. Elevation of serum lactate dehydrogenase in AL amyloidosis reflects tissue damage and is an adverse prognostic marker in patients not eligible for stem cell transplantation. Br J Haematol. 2017;178:888-95.
5. Farhana A, Lappin SL. Biochemistry, lactate dehydrogenase. 2023 May 1. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan.
6. Fang Y, Zhang Y, Zhang X. The elevated lactate dehydrogenase to albumin ratio is a risk factor for developing sepsis-associated acute kidney injury: a single-center retrospective study. BMC Nephrol. 2024;25:201.
7. Deng Y, Li X, Lai Q, Wang F, Zhang C, Yang Y, et al. Prognostic implication of lactic dehydrogenase-to-albumin ratio in critically ill patients with acute kidney injury. Clin Exp Nephrol. 2023;27:349-57.
8. Korzets Z, Plotkin E, Bernheim J, Zissin R. The clinical spectrum of acute renal infarction. Isr Med Assoc J. 2002;4:781-4.
9. Hazanov N, Somin M, Attali M, Beilinson N, Thaler M, Mouallem M, et al. Acute renal embolism. Forty-four cases of renal infarction in patients with atrial fibrillation. Medicine (Baltimore). 2004;83:292-9.
10. Ahn SB, Lee JY. Clinical differentiation between acute renal infarction and acute ureteral stone in the emergency department: a single-center retrospective case-control study. Am J Emerg Med. 2021;50:322-9.
11. Bourgault M, Grimbert P, Verret C, Pourrat J, Herody M, Halimi JM, et al. Acute renal infarction: a case series. Clin J Am Soc Nephrol. 2013;8:392-8.
12. Sert ET, Kokulu K. Relationship between microscopic haematuria and hydronephrosis in urolithiasis. Int J Clin Pract. 2021;75:e14688.
13. Mutailifu S, Zhu Q, Wang M, Zhang D, Song S, Li N. Association between lactate dehydrogenase/albumin ratio and in-hospital mortality in patients with acute aortic dissection. J Inflamm Res. 2025;18:6281-92.
14. Colgan SM, Mukherjee S, Major P. Hypoxia-induced lactate dehydrogenase expression and tumor angiogenesis. Clin Colorectal Cancer. 2007;6:442-6.
15. Fort J, Segarra A, Matas M, Segarra A, Camps J. Renal artery embolism: prospective study of 41 patients based on a diagnostic and therapeutic algorithm. The Open Urology & Nephrology Journal al. 2008;1:9-15.
16. Eltawansy SA, Patel S, Rao M, Hassanien S, ManiarM. Acute renal infarction presenting with acute abdominal pain secondary to newly discovered atrial fibrillation: a case report and literature review. Case Rep Emerg Med. 2014;2014:981409.
17. AntopolskyM, Simanovsky N, Stalnikowicz R, Salameh S, Hiller N. Renal infarction in the ED: 10-year experience and review of the literature. Am J Emerg Med. 2012;30:1055-60.
18. Nah S, Han S, Kim HB, Chun S, Kim S, Woo S, et al. Predictors of renal infarction in patients presenting to the emergency department with flank pain: a retrospective observational study. PLoS One. 2021;16:e0261054.
19. Zhang ZG, Liu XM. Clinical characteristics of patients with acute renal infarction: an analysis of 52 patients in a single center. Beijing Da Xue Xue Bao Yi Xue Ban. 2019;51:863-9.
20. Yang J, Lee JY, Na YJ, Lim SY, Kim MG, Jo SK, et al. Risk factors and outcomes of acute renal infarction. Kidney Res Clin Pract. 2016;35:90-5.
21. Sheinenzon A, Shehadeh M, Michelis R, Shaoul E, Ronen O. Serum albumin levels and inflammation. Int J Biol Macromol. 2021;184:857-62.
22. Bai Z, Bernardi M, Yoshida EM, Li H, Guo X, Méndez-Sánchez N, et al. Albumin infusion may decrease the incidence and severity of overt hepatic encephalopathy in liver cirrhosis. Aging. 2019;11:8502-25.
23. Özgür Y, Akın S, Yılmaz NG, Gücün M, Keskin Ö. Uric acid albumin ratio as a predictive marker of short-term mortality in patients with acute kidney injury. Clin Exp Emerg Med. 2021;8:82-8.
24. Garcia-Martinez R, Caraceni P, Bernardi M, Gines P, Arroyo V, Jalan R. Albumin: pathophysiologic basis of its role in the treatment of cirrhosis and its complications. Hepatology. 2013;58:1836-46.
25. Zoanni B, Brioschi M, Mallia A, Gianazza E, Eligini S, Carini M, et al. Novel insights about albumin in cardiovascular diseases: focus on heart failure. Mass Spectrom Rev. 2023;42:1113-28.
26. Sert ZS, Yılmaz SA, Seçilmiş Ö, Abuşoğlu S, Ünlü A, Çelik Ç. Effect of calcium and vitamin D supplementation on the clinical, hormonal, and metabolic profile in non-obese women with polycystic ovary syndrome. Ir J Med Sci. 2022;191:2657-62.
27. Levitt DG, Levitt MD. Human serum albumin homeostasis: a new look at the roles of synthesis, catabolism, renal and gastrointestinal excretion, and the clinical value of serum albumin measurements. Int J Gen Med. 2016;9:229-55.
28. Arques S. Serum albumin and cardiovascular disease: state-of-the-art review. Ann Cardiol Angeiol. 2020;69:192-200.
29. Chu M, Niu H, Yang N, Wang D, Liu Y, Mao X, et al. High serum lactate dehydrogenase to albumin ratio is associated with increased risk of poor prognosis after ischemic stroke. Clin Neurol Neurosurg. 2024;237:108120.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.72677
Eurasian J Emerg Med. 2026;25: 228-36

Machine Learning vs. Conventional Statistics in Predicting Medical Outcomes: A Comparative Performance Analysis Using a Viral Disease Dataset

Fatih İkiz[1], Ahmet Ak[2]

[1]University of Health Sciences Türkiye, Konya Beyhekim Training and Research Hospital, Clinic of Emergency Medicine, Konya, Türkiye
[2]Selçuk University Faculty of Medicine, Department of Emergency Medicine, Konya, Türkiye

Abstract

Aim: Accurate mortality prediction is fundamental for clinical resource allocation and personalized patient management across diverse medical conditions in patients admitted to emergency department. This study evaluates the predictive capacity of laboratory and clinical variables by conducting a direct comparison between conventional statistical models and advanced machine learning (ML) algorithms.

Materials and Methods: Mortality-associated variables were first analyzed using conventional approaches, including basic comparative tests, logistic regression, and ROC analyses. Subsequently, eleven ML algorithms were deployed to benchmark their performance against these conventional methods. The dataset was partitioned into a 75% training set and a 25% testing set. Models were evaluated based on sensitivity (recall) , area under the curve (AUC), and overall accuracy. Features of importance were defined as ranks assigned by the most robust model to identify key clinical predictors.

Results: While conventional statistical methods achieved a maximum sensitivity of 81.2% (AUC=0.906), ML algorithms significantly outperformed conventional statistical methods. Among the eleven algorithms, BayesNet emerged as the superior model, with a sensitivity of 88.9% and an overall classification accuracy of 92.3%. The analysis demonstrates that ML techniques capture complex, non-linear interactions within clinical data that standard logistic regression may overlook.

Conclusion: ML algorithms offer a substantial improvement in predictive performance over conventional statistical methods for mortality prediction in patients with complex viral diseases admitted to the emergency department. These findings suggest that integrating ML into clinical decision-support systems can provide more precise risk stratification than traditional prognostic tools.

Keywords: Artificial intelligence, machine learning, mortality, logistic regression, BayesNet

Introduction

The complex, multifactorial nature of viral diseases such as coronavirus disease 2019 (COVID-19), involving intricate interactions between numerous clinical and laboratory variables and potentially exhibiting non-linear relationships, may inherently limit the predictive capacity of these conventional, linear-based statistical approaches. Statistical methods, although their analytical forms can be more readily controlled and analyzed, depend on assumptions for their feasibility. Limitations may arise during the evaluation of numerous independent variables and multifactorial assessments (1).

In recent years, the field of machine learning (ML) has witnessed remarkable advancements, offering a suite of sophisticated algorithms capable of discerning complex patterns and making accurate predictions from high-dimensional datasets. In the realm of medicine, ML techniques such as decision trees, random forests, gradient boosting machines, support vector machines, and artificial neural networks have demonstrated considerable promise in various diagnostic and prognostic applications (2). Their ability to capture complex relationships, automatically identify feature interactions, and handle large volumes of data presents a compelling alternative and potential improvement

Corresponding Author: Fatih İkiz MD, University of Health Sciences Türkiye, Konya Beyhekim Training and Research Hospital, Clinic of Emergency Medicine, Konya, Türkiye
E-mail: sultanmehmet01@hotmail.com **ORCID ID:** orcid.org/0000-0003-3710-2193

Cite this article as: İkiz F, Ak A. Machine learning vs. conventional statistics in predicting medical outcomes: a comparative performance analysis using a viral disease dataset. Eurasian J Emerg Med. 2026;25: 228-36.

Received: 02.01.2026
Accepted: 22.03.2026
Published: 30.03.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

over conventional statistical methods, which still require further investigation. A comprehensive comparison focusing on the synergistic predictive value of coagulation parameters combined with standard clinical variables remains an area for further investigation.

Because it is a multifaceted and multifactorial disease and the effects of coagulation parameters on clinical outcomes have been demonstrated, the COVID-19 dataset was selected for use in this study. The dataset was deemed suitable for our comparison model for these reasons: it contains a sufficient sample size and well-defined clinical and laboratory outputs. To our knowledge, this is the first study to compare ML implementations with traditional statistical methods in this way for any viral disease, including COVID-19. The number of studies investigating the primary success rates of ML and statistical methods in medicine is limited. In this respect, we believe that our study will make a significant contribution to the literature and offer important insights to researchers regarding methodology selection.

Materials and Methods

Our study was conducted in the emergency department and focused on prognostic outcomes of patients, including their initial admission data. The study protocol was designed in accordance with the ethical principles of the Declaration of Helsinki and was approved by the Selçuk University Rectorate Local Ethics Committee of a university hospital (approval no: 2025/218, date: 08.04.2025). The data between 19.03.2020 and 26.10.2020 were accessed retrospectively with permission. Therefore, following ethical approval, data were retrospectively accessed between 08.04.2025 and 08.05.2025 and were obtained from patients aged 18 years and older who were admitted to the university hospital's emergency department and diagnosed with COVID-19 between 19.03.2020 and 26.10.2020. The authors had access to medical records that could identify individual participants during data collection. Because the study is retrospective and the data were provided by the hospital anonymously (with no names and confidential info), no consent was obtained in this regard. The research was concluded on 01.06.2025.

Research Model

This retrospective case-control study was conducted using data from 620 patients diagnosed with COVID-19 during the pandemic period between 19.03.2020 and 26.10.2020. Our study utilized coagulation markers to investigate parameters associated with 30-day in-hospital mortality [D-dimer, fibrinogen, prothrombin time (PT), activated PT (aPTT), and platelet]. Clinical findings glasgow coma scale (GCS), comorbidities, computerized tomography (CT) findings, and vital parameters at first admission were also assessed, which may be related to mortality. PCR positivity has been accepted as a criterion for COVID-19 diagnosis in our study. Ground-glass opacities, consolidations, and paving patterns are considered positive CT findings suggestive of COVID-19. The relevant findings were investigated using statistical methods and ML models, and the results were evaluated by a comprehensive comparison (conventional statistics vs. ML) (Figure 1).

Statistical Analysis

Statistical analyses were performed using SPSS 27.0 (IBM Inc., Chicago, IL, USA). The Kolmogorov-Smirnov test, histogram analyses, skewness and kurtosis, and Q-Q plots were used to evaluate the normality of continuous variables. Continuous variables were expressed as interquartile ranges [median (minimum-maximum)] or mean ± standard deviation. Qualitative data are expressed as frequencies (N) and percentages (%). Relationships of continuous variables between two groups were examined using the Mann-Whitney U test or the independent-samples t-test. The effect profiles of variables on mortality were evaluated using univariate and multivariate logistic regression analyses, limited to 10 events per variable. Box-Tidwell and multicollinearity assumptions were assessed prior to multivariate logistic regression analysis. Relationships between qualitative variables were investigated using Pearson's chi-square test or Fisher's exact test. Throughout the study, the type I error rate was set at 5% (α=0.05), and p-values less than 0.05 were considered statistically significant.

ML Methods

The survival and non-survival groups were separated to ensure homogeneous class distributions in the training and test sets. Subsequently, the sample was randomly divided into two independent groups: a 75% training set and a 25% test set. The data used in the test set were not used in the training set. To ensure that the models were trained and evaluated on representative samples, this split was performed using stratified sampling based on the binary outcome variable (survival/non-survival groups). This stratification ensured that the ratio of survival to non-survival classes was maintained in both the training and test sets. The test set was strictly held out and used only once for the final unbiased evaluation of the best model selected. Model development and initial comparison were conducted entirely on the 75% training set using 5-fold stratified cross-validation.

ML models and data mining were carried out and evaluated using WEKA software, version 3.8.6. ML models were run on a Windows 11 operating system using an Intel i7 CPU, 16 GB of RAM, and an NVIDIA GTX 1660 Ti 8 GB graphics card. The selection of ML algorithms was designed to represent a broad spectrum of computational approaches, including Bayesian

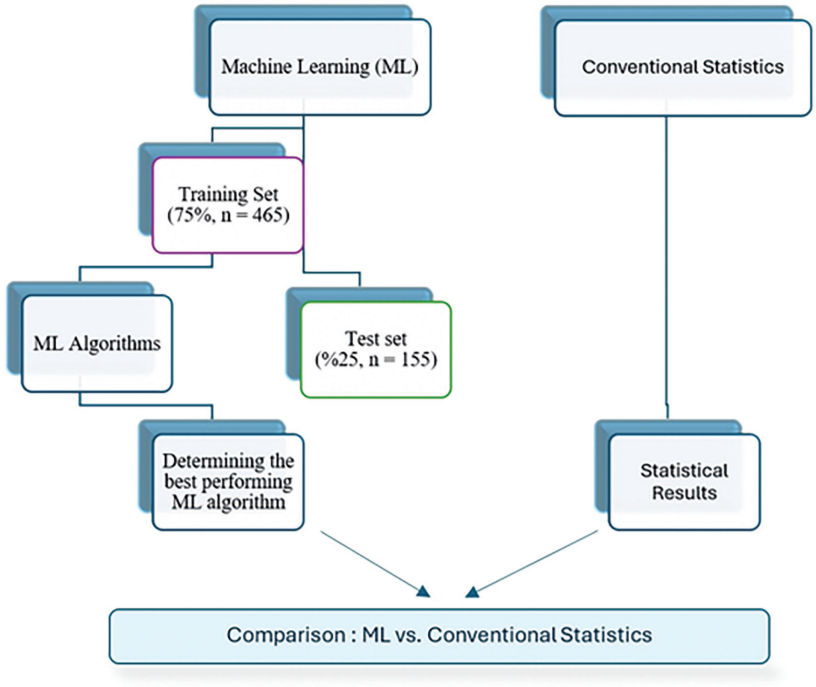


Figure 1. General flow chart of the study
ML: Machine learning

networks, decision trees, rule-based classifiers, and functional models. This diversity ensures that the comparison accounts for both linear and non-linear data structures. Furthermore, these specific algorithms were chosen based on their established performance in the medical prognostic literature and their availability within the WEKA 3.8.6 framework, thereby allowing standardized, reproducible benchmarking against conventional statistical methods. The 11 most popular ML models suitable for our study's model and dataset were applied and reported. To evaluate the predictive capabilities of the ML algorithms, mean absolute error, root mean squared error, correlation coefficients, Matthews correlation coefficient, F1 score, and recall were reported. The most successful ML algorithm was selected based on achieving the highest area under the curve (AUC) and F1-score. The AUC was chosen as the primary metric to select the model with the greatest overall discriminatory power, and the F1-score-the harmonic mean of sensitivity (recall) and precision-was used as a secondary measure to confirm a clinically relevant balance between minimizing false negatives and false positive (FP). In cases where a clear trade-off existed between the two metrics, the algorithm with the higher AUC was prioritized. Metrics such as true positive (TP) rate, FP rate, precision, and AUC were assessed for clinical utility. Among the algorithms with acceptable overall accuracy (>90%) and F1 scores (>0.70), the algorithm with the highest AUC value-also taking into account clinical adaptability and value in the medical field-was deemed the most suitable when evaluated from these perspectives. In this regard, it was concluded that the BayesNet algorithm was optimal. For the BayesNet, metrics such as the direction of use of the relevant attributes (variables) and feature importance in their prediction of the relevant class were reported as ranks. The structure of the final BayesNet was determined using the K2 algorithm with a batch size of 100, and the feature importance and direction were determined by analyzing the conditional probability tables (CPTs). The CPTs were estimated using the SimpleEstimator function (weka.classifiers.bayes.net.estimate.SimpleEstimator) with a Laplace correction (alpha=0.5). After CPT estimation, the features were ranked based on the magnitudes of the log-odds ratios. Subsequently, derived ranks are grouped as low (0-25th percentile), moderate (25th-75th percentile), and high (>75th percentile) based on the interquartile range. Missing values were handled based on the classifier type. For non-tree-based models, missing values were imputed prior to model training using the median of the training data for numerical features and the mode for categorical features. For decision tree and rule-based classifiers, missing values were handled internally by the classifier's default method, which distributes instances with unknown values fractionally across the branches of the decision structure based on the observed training data distribution.

Results

Age, gender, vital signs, GCS, coagulation parameters, positive CT findings, smoking history, and comorbidities were compared between survival groups. The results revealed that age, heart rate, respiratory rate, fibrinogen, international normalized ratio, and PT were significantly higher in non-survivors ($p<0.001$ for all). On the other hand, blood oxygen saturation (SpO_2), systolic blood pressure (SBP), diastolic blood pressure (DBP), GCS, and aPTT values were significantly higher in the survival group ($p<0.001$ for all), while no significant difference was observed between the groups in terms of gender and platelet levels ($p>0.05$). When comorbidities were examined, the prevalence of diabetes mellitus, hypertension, hyperlipidemia, congestive heart failure, vascular disease, malignancy, chronic obstructive pulmonary disease, chronic kidney disease, and stroke was higher in the non-survival group than in the survival group (Table 1).

Age, gender, coagulation parameters, comorbidity, and CT positivity were examined using univariate logistic regression analysis. Increases in age ($p<0.001$), D-dimer ($p<0.001$), fibrinogen ($p<0.001$), and PT levels ($p=0.009$), as well as decreases

Table 1. Comparison of the age, clinical and laboratory variables among survival groups

Variables		Survival (n=551, 88.9%)	Non-survival (n=69, 11.1%)	Overall	p
	Distribution†				
Age (years)		43.3±16.2	68.3±17.9	46.1±18.2	**<0.001**[a]
Hear rate (min^{-1})		82 (64-150)	82 (71-152)	82 (64-152)	**<0.001**[b]
Respiratory rate (min^{-1})		16 (12-40)	16 (14-40)	16 (12-40)	**<0.001**[b]
SpO_2 (%)		97 (8-99)	83 (45-99)	96 (8-99)	**<0.001**[b]
SBP (mmHg)		120 (100-180)	120 (75-155)	120 (75-180)	**<0.001**[b]
DBP (mmHg)		75 (50-110)	70 (50-90)	75 (50-110)	**<0.001**[b]
GCS		15 (14-15)	15 (3-15)	15 (3-15)	**<0.001**[b]
D-dimer (ng/mL)		307 (80-6610)	1173 (173-9935)	326 (80-9935)	**<0.001**[b]
Fibrinogen (ng/dL)		335 (45.1-1222)	499 (79-1017)	341 (45.1-1222)	**<0.001**[b]
INR		1 (0.04-24.3)	1.04 (0.87-9)	1 (0.04-24.3)	**<0.001**[b]
PT (sec)		11.6 (9.4-40.1)	12.4 (1.4-100)	11.6 (1.4-100)	**<0.001**[b]
aPTT (sec)		28 (16-60)	26 (16-39.5)	28 (16-60)	**<0.001**[b]
Platelet (K/µL)		201 (40-684)	198 (25-457)	201 (25-684)	0.418[b]
CT lesions in favor of COVID-19		280 (50.8%)	57 (82.6%)	337 (54.4%)	**<0.001**[c]
Gender	Female	253 (45.9%)	28 (40.6%)	281 (45.3%)	0.401[c]
	Male	298 (54.1%)	41 (59.4%)	339 (54.7%)	
DM	Yes	38 (6.9%)	21 (30.43%)	59 (9.52%)	**<0.001**[c]
HT	Yes	60 (10.89%)	31 (44.93%)	91 (14.68%)	**<0.001**[c]
HL	Yes	13 (2.36%)	9 (13.04%)	22 (3.55%)	**<0.001**[d]
Smoking	Yes	55 (9.98%)	8 (11.59%)	63 (10.16%)	0.676[c]
CHF	Yes	10 (1.81%)	10 (14.49%)	20 (3.23%)	**<0.001**[d]
Vascular disease	Yes	19 (3.45%)	16 (23.19%)	35 (5.65%)	**<0.001**[d]
Malignancy	Yes	9 (1.63%)	9 (13.04%)	18 (2.9%)	**<0.001**[d]
COPD	Yes	12 (2.18%)	18 (26.09%)	30 (4.84%)	**<0.001**[d]
CKD	Yes	10 (1.81%)	12 (17.39%)	22 (3.55%)	**<0.001**[d]
Stroke	Yes	2 (0.36%)	3 (4.35%)	5 (0.81%)	**0.011**[d]
Hepatic disease	Yes	4 (0.73%)	0 (0%)	4 (0.65%)	>0.99[d]
Autoimmune disease	Yes	3 (0.54%)	2 (2.9%)	5 (0.81%)	0.098[d]

†: Data are expressed as interquartile ranges (median, minimum-maximum) or mean ± SD, ǂ: Ground glass opacities, consolidations, paving patterns, [a]: Independent t-test, [b]: Mann-Whitney U test, [c]: Pearson chi-square test, [d]: Fisher's exact test SD: Standard deviation, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, GCS: Glascow coma score, INR: International normalised ratio, PT: Prothrombin time, aPTT: Activated partial thromboplastin time, CKD: Chronic kidney disease, DM: Diabetes mellitus, HT: Hypertension, HL: Hyperlipidemia, CHF: Congestive heart failure, COPD: Chronic obstructive pulmonary disease, CT: Computerized tomography, COVID-19: Coronavirus disease 2019, SpO_2: Blood oxygen saturation

in aPTT levels (p<0.001), were associated with higher mortality. Additionally, comorbidities were dichotomized into binary groups, and patients with comorbidities had an 8.61-fold higher mortality risk compared to those with without comorbidity [odds ratio (OR)=8.61, p<0.001]. Similarly, the mortality risk increased 4.6-fold among patients with positive CT results (OR=4.6, p<0.001). Significant variables underwent multivariate logistic regression analysis. In the multivariate analysis, increased levels of D-dimer and advancing age continued to be associated with higher mortality rates (Table 2). GCS was excluded from these analyses because it violated the assumptions by exhibiting high multicollinearity.

ROC analysis was performed for age and coagulation parameters (D-dimer, fibrinogen, PT, and aPTT) that showed a significant association with mortality, and their predictive values for mortality were investigated. The examination revealed that at a cut-off age of 58.8 years, the sensitivity was 78.3% and the specificity was 82.6%. Among the laboratory parameters, the highest AUC value was observed in the D-dimer at a cut-off value of ≥638.5 ng/mL (AUC=0.832, 75.8% sensitivity and 83.1% specificity, p<0.001). This was followed by fibrinogen, PT, and aPTT. The ROC analysis data for these variables are summarized in Table 3 and Figure 2.

A systematic assessment of combinations of variables and their predictive capabilities for mortality has been statistically analyzed. In step 1, the combination of age and comorbidity demonstrated a sensitivity of 81.2% and a specificity of 80.6% (AUC=0.848, p<0.001). In step 2, the inclusion of vital signs into the prior combination demonstrated a sensitivity of 80.9% and an improved specificity of 90.6% (AUC=0.905, p<0.001). At step 3, age, comorbidities, vital signs, and laboratory data, which were shown to have a significant impact on outcomes (D-dimer, fibrinogen, PT, and aPTT), are integrated. Although sensitivity remains relatively unchanged at 80.4%, specificity has risen to 92.9% (AUC=0.905, p<0.001). In the final stage, the clinical data (GCS and CT results) are combined, demonstrating that no substantial change is observed compared with step 3. The maximum sensitivity for mortality was 81.2%, with a specificity of 92.9%. In line with the results, considering that vital signs significantly increase specificity, whereas other combinations do not significantly increase sensitivity or specificity, vital signs are among the parameters with the greatest impact on mortality and contribute most to specificity (Table 4).

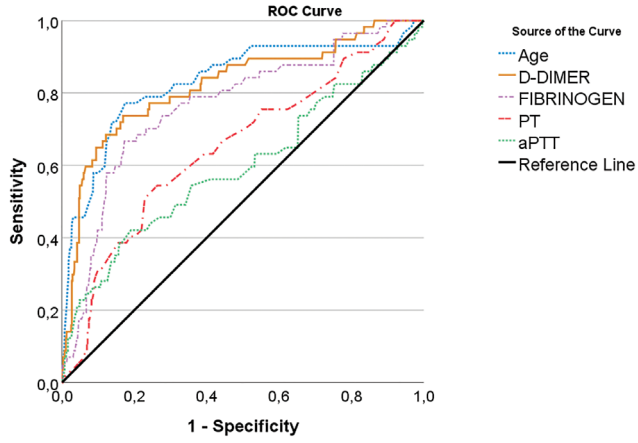


Figure 2. ROC analysis graph of variables in predicting mortality

PT: Prothrombin time, aPTT: Activated partial thromboplastin time

Table 2. Univariate and multivariate logistic regression tests for mortality

Variable	Univariate LR				Multivariate LR¶	
	B	Nagelkerke R^2	P	OR (95% CI)	OR	P
D-dimer (ng/mL)	0.001	0.235	**<0.001**	1.001 (1.00-1.002)	1.00	**0.029**
Fibrinogen (ng/dL)	0.006	0.148	**<0.001**	1.006 (1.004-1.007)	1.002	0.103
INR	0.079	0.003	0.328	1.083 (0.924-1.269)	-	-
PT (sec)	0.112	0.037	**0.009**	1.119 (1.028-1.217)	1.061	0.221
APTT (sec)	-0.135	0.062	**<0.001**	0.873 (0.822-0.928)	0.944	0.111
Platelet (K/µL)	-0.002	0.003	0.360	0.998 (0.995-1.002)	-	-
Age (years)	0.081	0.333	**<0.001**	1.084 (1.065-1.103)	1054	**<0.001**
Gender	0.218	0.002	0.402	1.243 (0.747-2.068)	-	-
Comorbidity (overall)	2.154	0.191	**<0.001**	8.61 (4.78-15.54)	1.853	0.113
CT lesions in favor of COVID-19	1.525	0.086	**<0.001**	4.60 (2.413-8.758)	1.266	0.560

¶: Box-Tidwell assumptions have been met and variables with high multicollinearity was excluded in the model (Nagelkerke R^2 of the multivariate model= 0.371). The model is built in line with the events per variable ~10 rule, COVID-19: Coronavirus disease 2019, OR: Odds ratio, CI: Confidence interval, LR: Logistic regression, INR: International normalised ratio, PT: Prothrombin time, aPTT: Activated partial thromboplastin time, CT: Computerized tomography

Table 3. ROC Analysis of parameters in favor of mortality

	AUC	95% CI		Cut-off ¶	Sensitivity (%)	Specificity (%)	P
		Lower limit	Upper limit				
D-dimer (ng/mL)	0.832	0.772	0.892	≥638.5	75.8	83.1	**<0.001**
Fibrinogen (ng/dL)	0.762	0.695	0.830	≥443.5	67.7	81.7	**<0.001**
PT (sec)	0.658	0.585	0.731	≥12.35	52.2	77.1	**<0.001**
aPTT (sec) †	0.631	0.550	0.712	≤24.85	43.5	84.0	**<0.001**
Age (years)	0.841	0.784	0.898	≥58.8	78.3	82.6	**<0.001**

†: Cut-off values determined based on Youden J index, ¶: Lower values are associated with positive (non-survival) results AUC: Area under the curve, CI: Confidence interval, PT: Prothrombin time, aPTT: Activated partial thromboplastin time

In our study, ML models established using 11 algorithms based on a concept suitable for our data structure were compared. The examination showed that the algorithms with the highest precision values were logistic, simple logistic, support vector machine, and multilayer perceptron. The highest overall accuracy (for combined classification of survival and non-survival groups) was achieved by the J48 algorithm. On the other hand, the best-performing algorithms in mortality prediction were BayesNet and LogitBoost. The performances of BayesNet and LogitBoost algorithms were evaluated together with F1 score, Matthew's correlation coefficient (MCC), mean absolute error, root mean squared error, overall accuracy, Kappa (κ) and AUC values, and it was shown that the best performing algorithm was BayesNet (recall= 88.9%, AUC=88.6%, F1 score= 0.727, MCC=0.70, overall accuracy= 92.3%, mean absolute error= 0.08, κ=0.704) (Figure 3). As a result, the algorithms exhibited different predictive capabilities across different aspects, resulting in variation in recall (sensitivity), TP rate, and AUC values. The prediction features of all algorithms are summarized in Table 5.

After determining the best algorithm, the priority variables used by the algorithm and their rank scores were identified.

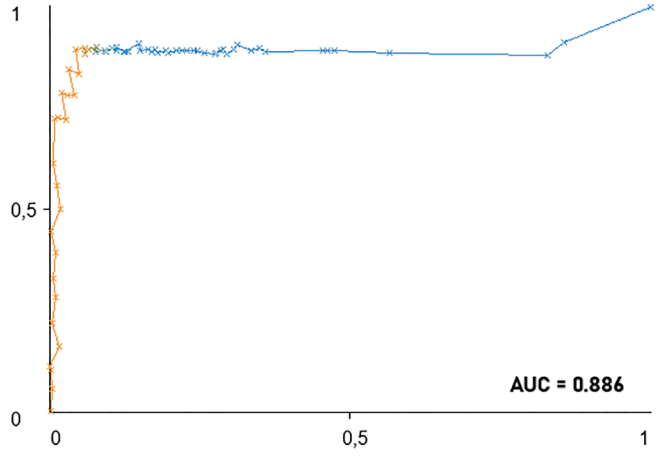


Figure 3. ROC curve of the best ML model with highest precision (BayesNet)

ML: Machine learning, AUC: Area under the curve

Thus, a ranking score was obtained for the variables used by the algorithm to determine the target class, allowing identification of variables that ranked highest in the prediction and of those with low or no effect. In the evaluation, it was understood that the most valuable parameters in mortality prediction (high-impact group) were GCS, DBP, SBP, D-dimer, SpO_2, and respiratory rate, respectively. On the other hand, the most ineffective variable, never used in prediction, was platelet level (rank score= 0). Similarly, the impact of smoking history, gender, history of hepatic disease, positive CT findings, heart rate, and history of autoimmune disease was noted to be low. The remaining attributes had moderate impacts on the prediction of the target class. The attributes and rank scores of all variables are summarized in Figure 4.

Discussion

Our study compared conventional statistical methods with ML algorithms for predicting COVID-19 mortality, showcasing the potential of ML in this complex clinical scenario. We meticulously evaluated eleven ML models, with the BayesNet algorithm demonstrating the best performance in predicting mortality (recall= 88.9%, AUC=88.6%). The J48 algorithm, although not the top performer for mortality alone, achieved the highest overall accuracy in classifying both survival and mortality groups. Importantly, the ML approach identified GCS, DBP, SBP, D-dimer, SpO_2, and respiratory rate as the most influential parameters for mortality prediction, while platelet levels were found to be least impactful (with a rank score of 0). The implementation of ML in medicine extends far beyond COVID-19 prognosis. ML algorithms offer the capability to analyze vast and complex datasets, identify non-linear relationships, and discover novel patterns that might be missed by conventional statistical methods. This leads to enhanced diagnostic accuracy, improved operational workflows, robust clinical decision support, and ultimately, better patient outcomes (3,4). As highlighted in recent reviews, ML is reshaping fields like pathology through automated image analysis, biomarker discovery, and drug development (3). The

Table 4. Stepwise evaluation of combined variables and their capabilities in predicting mortality

Combination of variables	AUC (95% CI)	p-value	Sensitivity (%)†	Specificity (%)†
Step 1: Age, comorbidity (overall)	0.848 (0.791-0.904)	**<0.001**	81.2%	80.6%
Step 2: Age, comorbidity (overall), vital signs[a]	0.905 (0.859-0.952)	**<0.001**	80.9%	90.2%
Step 3: Age, comorbidity (overall), vital signs and laboratory variables[b]	0.905 (0.851-0.958)	**<0.001**	80.4%	92.9%
Step 4: Age, comorbidity (overall), vital signs, laboratory variables and clinical data[c]	0.906 (0.852-0.959)	**<0.001**	80.4%	92.9%

[a]: Heart rate, respiratory rate, SBP, DBP, SpO_2
[b]: D-dimer, Fibrinogen, PT and aPTT
[c]: GCS, CT lesions in favor of COVID-19
†: Best cut-off values is determined in accordance with Youden J index
AUC: Area under the curve, CI: Confidence interval, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, SpO_2: Blood oxygen saturation, PT: Prothrombin time, aPTT: Activated partial thromboplastin time, GCS: Glascow coma score, CT: Computerized tomography, COVID-19: Coronavirus disease 2019

Table 5. Evaluation of ML success and prediction capabilities for each model

	Evaluation of model success										
	Model		ML statistics - predicting mortality								
ML algorithms*	Mean absolute error	Root mean squared error	Kappa (κ)	TP rate	FP rate	Precision	Recall	Overall accuracy¥	F1 score	MCC	ROC (AUC)
J48	0.094	0.236	0.723	0.778	0.036	0.737	0.778	94.2	0.757	0.724	0.842
Logistic	0.074	0.224	0.586	0.444	0.0	1.00	0.444	93.6	0.615	0.644	0.868
Simple logistic	0.08	0.213	0.530	0.389	0.0	1.00	0.389	92.9	0.560	0.60	0.880
Random forestΦ	0.186	0.233	0.736	0.722	0.022	0.813	0.722	94.8	0.765	0.737	0.861
BayesNet†	0.08	0.259	0.704	0.889	0.073	0.615	0.889	92.3	0.727	0.700	0.886
SMO	0.07	0.266	0.529	0.389	0.0	1.00	0.389	92.9	0.560	0.600	0.694
MultilayerPerceptron	0.08	0.267	0.443	0.333	0.007	0.857	0.333	91.6	0.480	0.503	0.884
Random tree	0.297	0.545	0.206	0.667	0.292	0.231	0.667	70.3	0.343	0.254	0.687
Decision stump	0.133	0.247	0.631	0.833	0.08	0.577	0.833	91.0	0.682	0.646	0.877
Rep tree	0.11	0.237	0.652	0.611	0.022	0.786	0.611	93.6	0.688	0.659	0.878
PART decision	0.156	0.364	0.423	0.722	0.146	0.394	0.722	83.9	0.510	0.451	0.797
LogitBoost	0.296	0.389	0.577	0.889	0.117	0.50	0.889	88.4	0.640	0.611	0.891

*: Training and test split: 75.0%, 25.0%
¥: In predicting both groups (mortality and survival)
†: The best-performing algorithm for predicting mortality classified 16 patients (out of 18)
Φ: The best-performing for overall prediction correctly classified 147 patients (out of 155)
TP: True positive, FP: False positive, MCC: Matthew's correlation coefficient, AUC: Area under the curve, SMO: Support vector machine

core strength of ML lies in its ability to learn from data, making it a powerful tool for personalized medicine, risk stratification, and early disease detection (5).

Our findings indicate that, while conventional methods, particularly stepwise logistic regression incorporating age, comorbidities, vital signs, and key laboratory data (including coagulation factors), achieved good specificity (92.9%) and sensitivity (80.4%), the BayesNet ML model achieved higher recall (sensitivity) for mortality prediction (88.9%) and a comparable AUC (88.6%). This observation is echoed in the wider literature, where ML models often demonstrate enhanced predictive performance compared to conventional statistical models, although this is not universally the case and is context dependent (6). For instance, a systematic review and meta-analysis comparing ML with logistic regression for predicting outcomes after percutaneous coronary intervention found that ML models generally resulted in higher c-statistics, though the overall performance could be comparable, and many studies suffered from a high risk of bias (7). Another meta-analysis on predicting all-cause mortality in acute coronary syndrome patients showed that best-performing ML models had

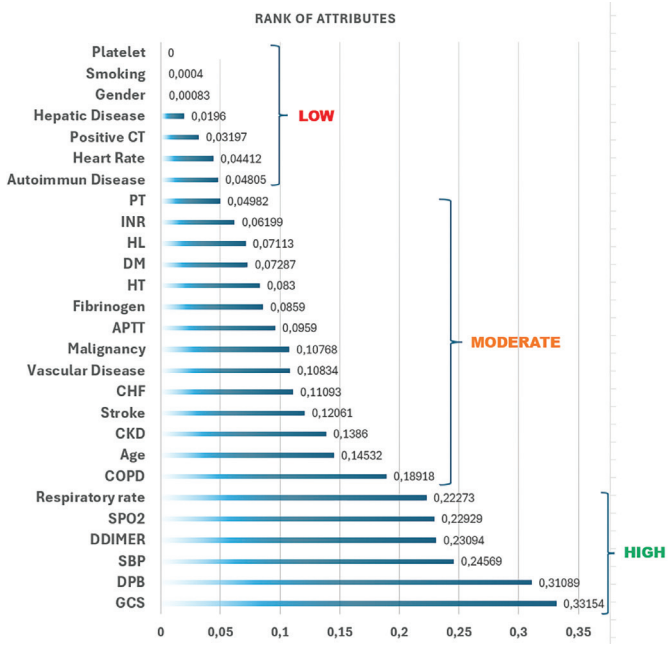


Figure 4. The worth of an attribute (feature of importance) with respect to mortality was evaluated. Ranks are statistically divided into quartiles: the top 25% is defined as "high," the middle 25-75% as "moderate," and the bottom 25% as "low"

CT: Computed tomography, PT: Prothrombin time, INR: International normalized ratio, HL: Hyperlipidemia, DM: Diabetes mellitus, HT: Hypertension, APTT: Activated partial thromboplastin time, CHF: Congestive heart failure, CKD: Chronic kidney disease, COPD: Chronic obstructive pulmonary disease, SpO_2: Blood oxygen saturation, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, GCS: Glasgow coma scale

a superior c-statistic (0.88) compared to conventional methods (0.82) (8). In a direct comparison between BayesNet and logistic regression, research indicates that BayesNet can be equally efficient (9) and, in certain cases, superior (10,11), particularly in managing complex variable interactions and integrating prior knowledge. The strength of conventional statistics, as used in this study, lies in its interpretability and its ability to readily analyze controlled data, though it relies on certain assumptions. ML models, while potentially more powerful in discerning complex patterns, can sometimes be perceived as "black boxes" making interpretation more challenging, a concern noted in the medical community (6). Using programming and statistical methods, this study addressed this issue by identifying the rank scores and feature importance used by the best-performing ML algorithm and by enhancing its interpretability.

This study demonstrates the prognostic significance of coagulation parameters, particularly D-dimer, for COVID-19 mortality, consistent with extensive external research. Furthermore, its comparative analysis provides valuable insights into the superior predictive performance of ML algorithms (e.g., BayesNet) compared with conventional statistical models (logistic regression) in this setting. The ability of ML to identify key predictive variables and handle complex data interactions positions it as a promising tool for improving risk stratification and guiding clinical decision-making in COVID-19 and other complex viral diseases. Future research should focus on external validation of these ML models in diverse patient cohorts to ensure generalizability. Continued efforts to enhance the interpretability of ML models and to integrate them seamlessly into clinical workflows will be crucial to realizing their full potential in advancing patient care. The combination of robust clinical understanding, sound statistical principles, and advanced computational techniques, as exemplified in this work, paves the way for more accurate and timely medical predictions.

While this study utilized a COVID-19 dataset due to its clinical richness and the heterogeneous nature of the disease, the findings offer broader insights into methodology selection for predicting outcomes in similar clinical contexts. Our study not only emphasizes the need for other similar conceptual studies, but also serves as an ideal guide for the research and development of ML applications in the field of medicine.

Study Limitations

A limitation of our study is that the sample size of the output class (non-survival group) was inadequate to fit multivariate logistic regression models that included all variables. Therefore, we had to limit the events per variable rule to approximately 10, which required us to exclude some variables from the model. Additionally, unlike ML models, the statistical methods we employed required adherence to certain assumptions (e.g., Box-Tidwell test and multicollinearity). Consequently, we excluded variables that did not satisfy these assumptions. Another limitation of our study is the lack of external validation to assess the efficacy of the ML algorithms. The Nagelkerke R^2 value of 0.371 in the multivariate logistic regression model indicates that the current variables account for only 37.1% of the variance in mortality. This suggests the presence of many additional parameters that could be associated with mortality, highlighting the complex nature of COVID-19. These findings indicate that further studies to identify additional variables, with confounders defined may be necessary. Further assessment of ML algorithms and their comparisons with conventional statistical methods are required.

Conclusion

Conventional statistical methods require that certain assumptions be met and have limitations. With more algorithmic structures

and their ability to process complex datasets, ML algorithms can achieve higher sensitivity in mortality prediction. For complex, multifactorial diseases, ML methods are emerging as a promising approach.

Ethics

Ethics Committee Approval: The study protocol was designed in accordance with the ethical principles of the Declaration of Helsinki and was approved by the Selçuk University Rectorate Local Ethics Committee of a university hospital (approval no: 2025/218, date: 08.04.2025).

Informed Consent: Informed consent was waived due to the retrospective design.

Footnotes

Authorship Contributions

Concept: F.İ., A.A., Design: F.İ., Data Collection or Processing: F.İ., A.A., Analysis or Interpretation: F.İ., Literature Search: F.İ., Writing: F.İ.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Nørskov AK, Lange T, Nielsen EE, Gluud C, Winkel P, Beyersmann J, et al. Assessment of assumptions of statistical analysis methods in randomised clinical trials: the what and how. BMJ Evid Based Med. 2021;26:121-6. Epub 2020 Jan 27.
2. Ahsan MM, Luna SA, Siddique Z. Machine-learning-based disease diagnosis: a comprehensive review. Healthcare (Basel). 2022;10:541.
3. Hanna MG, Pantanowitz L, Dash R, Harrison JH, Deebajah M, Pantanowitz J, et al. Future of artificial intelligence-machine learning trends in pathology and medicine. Mod Pathol. 2025;38:100705. Epub 2025 Jan 4.
4. Deo RC. Machine learning in medicine. Circulation. 2015;132:1920-30.
5. Adlung L, Cohen Y, Mor U, Elinav E. Machine learning in clinical decision making. Med. 2021;2:642-65. Epub 2021 Apr 30.
6. Mann J, Lyons M, O'Rourke J, Davies S. Machine learning or traditional statistical methods for predictive modelling in perioperative medicine: a narrative review. J Clin Anesth. 2025;102:111782. Epub 2025 Feb 19.
7. Nayebirad S, Hassanzadeh A, Vahdani AM, Mohamadi A, Forghani S, Shafiee A, et al. Comparison of machine learning models with conventional statistical methods for prediction of percutaneous coronary intervention outcomes: a systematic review and meta-analysis. BMC Cardiovasc Disord. 2025;25:310.
8. Gupta AK, Mustafiz C, Mutahar D, Zaka A, Parvez R, Mridha N, et al. Machine learning vs traditional approaches to predict all-cause mortality for acute coronary syndrome: a systematic review and meta-analysis. Can J Cardiol. 2025;41:1564-83. Epub 2025 Feb 17.
9. Ducher M, Kalbacher E, Combarnous F, Finaz de Vilaine J, McGregor B, Fouque D, et al. Comparison of a Bayesian network with a logistic regression model to forecast IgA nephropathy. Biomed Res Int. 2013;2013:686150. Epub 2013 Nov 17.
10. Yuan B, Wang C, Fan Z, Liu C, Fang L, Ma L, et al. A Bayesian network-based approach for identifying risk factors and predicting ischemic stroke in infective endocarditis patients. Front Cardiovasc Med. 2024;10:1294229.
11. Ma L, Cai B, Qiao ML, Fan ZX, Fang LB, Wang CB, et al. Risk factors assessment and a Bayesian network model for predicting ischemic stroke in patients with cardiac myxoma. Front Cardiovasc Med. 2023;10:1128022.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.30676
Eurasian J Emerg Med. 2026;25: 237-45

Does Fear of Workplace Violence Reduce Life Satisfaction? The Mediating Role of Psychological Resilience Among Healthcare Workers

Mehtap Öztürk[1], Meliha Fındık[2]

[1]Selçuk University Faculty of Business and Economics, Department of Business Administration, Konya, Türkiye
[2]Balıkesir Üniversity Faculty of Medicine, Department of Emergency Medicine, Balıkesir, Türkiye

Abstract

Aim: This study investigates the relationship between fear of workplace violence and life satisfaction among healthcare workers and evaluates the potential mediating role of psychological resilience in this relationship.

Materials and Methods: A cross-sectional design was employed with 173 healthcare workers at Balıkesir University Faculty of Medicine Hospital, recruited via convenience sampling between March 15, 2024, and April 15, 2025. Data were collected using the Fear of Workplace Violence, Psychological Resilience, and Life Satisfaction scales. Mediation analysis was conducted using PROCESS Macro v5.0 (Model 4) with 5,000 bootstrap samples and a 95% confidence interval. Gender, age, professional experience, education, marital status, and prior exposure to violence were included as control variables.

Results: Approximately 24.3% of participants reported previous exposure to violence. Correlational analyses indicated that fear of workplace violence was significantly and negatively associated with both psychological resilience ($r=-0.382$, $p<0.001$) and life satisfaction ($r=-0.437$, $p<0.001$), whereas psychological resilience was positively associated with life satisfaction($r=0.262$, $p< 0.01$). In the mediation model, fear of violence significantly predicted lower psychological resilience B=-0.319) and had a strong direct adverse effect on life satisfaction (B=-0.517). The overall model explained 36.3% of the variance in life satisfaction ($R^2=0.363$).

Conclusion: Fear of workplace violence is a significant predictor of reduced life satisfaction; this association remains robust after controlling for sociodemographic factors and prior victimization. Notably, psychological resilience did not mediate this relationship within this sample. These findings suggest that individual-level interventions alone are insufficient; organizational strategies must prioritize reducing the perceived risk and fear of violence to protect healthcare workers' well-being.

Keywords: Workplace violence, fear of violence, psychological resilience, life satisfaction, healthcare workers, mediation analysis

Introduction

The field of healthcare services is one in which employees frequently encounter security threats and psychosocial stressors due to high workloads, shift work, and ongoing interactions with patients and their relatives. In this context, workplace violence is considered a significant occupational risk that affects employees' sense of security, functioning, and psychological well-being, extending beyond the incident itself. In explaining the negative consequences of violence, not only the presence of exposure but also the cognitive and affective processes following exposure are decisive; it is emphasized that the fear of experiencing violence in the workplace may be a central mechanism in revealing the effects of violence experiences on health and work outcomes (1). It has been reported that fear is a multidimensional experience comprising components of worry and perceived risk, and that it is a distinct stressor in environments characterized by intense, often unpredictable interactions, such as healthcare, where it is associated with psychological well-being (2).

Indeed, in the study by Demirbaş et al. (3), 87.8% of healthcare workers reported exposure to verbal violence and 17.8% to

Corresponding Author: Meliha Fındık MD, Balıkesir Üniversity Faculty of Medicine, Department of Emergency Medicine, Balıkesir, Türkiye
E-mail: meliha.findik@balikesir.edu.tr **ORCID ID:** orcid.org/0000-0001-8513-9842
Cite this article as: Öztürk M, Fındık M. Does fear of workplace violence reduce life satisfaction? The mediating role of psychological resilience among healthcare workers. Eurasian J Emerg Med. 2026;25: 237-45.

Received: 08.01.2026
Accepted: 22.03.2026
Published: 30.03.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

physical violence. It was shown that the perception that the spread of violent content on social media could increase violence was widespread, and that the "code white" application was widely used. These findings support the notion that the fear of violence (FV) should be considered not only as an individual exposure but also as a risk perception sustained by the institutional security climate and visible prevention/response mechanisms. Similarly, Tekin and Selcen (4) five-year review covering 2018-2022 indicates that attacks on healthcare services totaled 3,980 incidents, with 2,697 injuries and 1,095 deaths reported. The study also emphasized that the impact of attacks extends beyond the moment of the incident, disrupting access to healthcare services and leading to secondary losses; therefore, it drew attention to the need for well-structured and functional reporting/recording systems (4).

Life satisfaction (LS), one of the core cognitive indicators of subjective well-being, refers to an individual's assessment of their quality of life through a general and cognitive evaluation process (5,6). LS is a relatively stable outcome shaped by the interaction of internal resources and external conditions (7-10). Within this framework, the fear of experiencing violence, which creates a persistent perception of threat, is expected to negatively affect LS. Psychological resilience (PR), as a dynamic and developable adaptive capacity, can play a mediating and/or buffering role in stress-health processes (11-15).

However, in the literature, studies focusing on fear of future workplace violence (FFVW) have mainly concentrated on clinical distress indicators such as depressive symptoms, anxiety, sleep problems, and burnout. There is limited testing of its reflection on a more general subjective well-being outcome, such as LS, and of the mediating role of resilience in this relationship, in a comprehensive manner (15-18). Therefore, this study aims to examine the relationship between fear of workplace violence and LS among healthcare workers, and to test whether PR mediates this relationship.

Conceptual Framework

Fear of Workplace Violence

Fear of workplace violence refers to the cognitive and emotional response triggered by an employee's assessment of the likelihood of experiencing violence in the future, independent of any actual violent incident. This construct is considered one of the fundamental psychological processes explaining how exposure to violence translates into individual and organizational outcomes (1). Fear cannot be reduced to a mere momentary emotional state; it is defined as a multidimensional experience comprising worry about the future and perceived risk, which shape how employees interpret security threats in the work environment (2). Particularly in service contexts such as healthcare, where interaction intensity is high and unpredictability is pronounced, the "fear of FFVW" can function as a functional stressor, with significant effects on perceived safety and psychological well-being (18,19).

The level of FFVW is related both to individual exposure patterns and to objective and perceived risk conditions in the working context. In addition to direct exposure to physical or verbal violence, indirect forms of exposure, such as witnessing violence or learning about incidents involving colleagues, can also trigger similar fear responses and be associated with functional impairment (20,21). At the demographic and occupational level, findings indicate that female employees and younger staff report higher levels of fear; in the healthcare sector, both the risk of violence and fear are higher compared to industrial settings (16,20,22). Organizational conditions are also decisive: perceptions of organizational justice, adequacy of safety measures, and management support shape the intensity of fear. While perceived inadequacy of violence prevention policies may increase fear, perceived prevention and protective practices have been reported to reduce fear (23).

The analytical value of FFVW extends beyond a mere "consequence of exposure"; it also functions as a transmission mechanism mediating the effects of violence-related experiences on health and work outcomes. Early findings suggest that the fear and insecurity following exposure may impair mental health and performance more directly than the event itself (1). Current studies indicate that FFVW along with stress processes such as burnout in some models can partially or fully mediate the relationship between exposure to violence and depressive symptoms (15). Similarly, fear has been reported as one of the key explanatory pathways linking violence to reduced work ability (21). Furthermore, FFVW has been strongly associated with depressive symptoms, anxiety, and impaired sleep quality (17,18) and has been linked to burnout processes over time through the depletion of psychological resources under perceived threat (16). At the organizational level, it has been linked to decreased job satisfaction and commitment, increased intention to leave, and absenteeism/productivity loss, particularly in high-risk service environments (20,23). This body of evidence supports treating healthcare workers' FV not merely as a post-event reaction but as a risk indicator with comprehensive effects on well-being and functioning.

Psychological Resilience

Although PR has been treated as a relatively stable personality trait over time, the current literature increasingly defines it as a dynamic, developable, and context-sensitive adaptation process

(11,12). According to process-based approaches, resilience encompasses the capacity to maintain, sustain, or regain functionality by demonstrating positive adaptation in the face of meaningful adversity (13,14). Although the metaphor of "bouncing back" evokes recovery, contemporary frameworks view resilience not only as a return to the old equilibrium but also as the capacity to sustain purpose- and meaning-oriented functioning and to reconfigure adaptation in the face of prolonged adversity (24,25). This approach aligns with the view that resilience is a multi-system phenomenon shaped by the interaction of psychological, social, and biological systems over time, rather than a singular outcome (11). It is also argued that biological processes, including epigenetic mechanisms, can shape resilience in interaction with supportive and adverse contextual conditions (26).

When examining how resilience is strengthened, individual and environmental resources are seen to jointly contribute. Individual resources, such as self-efficacy, positive affect, LS, and sleep quality, have been consistently associated with resilience, particularly in professions characterized by high demands and uncertainty (27). At the mechanistic level, emotion regulation, developing tolerance to stressors, and structuring experiences into a coherent narrative (narrative sense-making) are key components of adaptation processes (28). At the social level, relational resources such as family functioning and social support are prominent in the development of resilience; it has been reported that family functioning can enhance resilience through LS, especially in collectivist cultures (29).

Resilience is considered not only a trait that enhances well-being but also a process that explains the effects of stressors. It has been reported that the negative impact of high-threat stressors, such as workplace violence, on depressive symptoms varies according to resilience and can explain interindividual differences (15). Similarly, resilience has been suggested to mediate the relationship between poor sleep quality and mental well-being (30) and to serve as a conduit mechanism in the reflection of psychological distress, such as anxiety, on behavioral outcomes (31). In the context of healthcare workers, this framework positions resilience as a psychological capacity that enables the maintenance of well-being under perceived threat and chronic stress.

Life Satisfaction

LS is an individual's general and cognitive assessment of their quality of life based on their own criteria and is one of the fundamental components of subjective well-being (5,6). This construct differs from more momentary emotional experiences, such as happiness, in that it reflects a relatively stable, long-term judgment of the harmony between desired life conditions and the current situation (32,33). Two main approaches are emphasized in explaining LS: the bottom-up perspective suggests the cumulative effect of satisfaction in areas such as work/family/health; the top-down perspective suggests that more stable tendencies, such as personality and general outlook, shape evaluations (7,8). Findings indicate that the interaction of internal resources, such as self-esteem, PR, and optimism, and external factors, such as social support, economic conditions, and health, shapes LS (9,10). The strong predictive power of self-esteem and the fact that this relationship can vary through intermediate processes such as body image or self-presentation tactics indicate that LS is fueled by the combined functioning of self-based cognitions and social-relational dynamics (10,34).

From a developmental perspective, LS may exhibit varying trends across age periods. It is noted that it declines with age during adolescence (35), while in older age, despite physical health decline, it can be maintained or even increased thanks to coping mechanisms and gerotranscendence processes; this is discussed as the "aging paradox" (36,37). PR has been reported to play a protective/restorative role in LS during stressful life events, illnesses, and crisis conditions such as pandemics (38,39), and social capital indicators (family functioning, friend support, social networks) have been reported to support LS (29,40). For healthcare workers, this body of research indicates that LS is an indicator of well-being that can become fragile under perceived threats; however, it can be sustained through protective psychological and social resources.

Research Hypotheses

Fear of workplace violence is defined as a multidimensional stress experience that reflects future threat perception, anxiety, and perceived risk, independent of exposure itself (1,2). Findings suggest that FFVW is associated with psychological well-being among healthcare workers (18,19) and classical evidence showing that fear acts as a pathway affecting mental health/functioning (1) suggest that this threat may also negatively affect the general cognitive assessment of quality of life. Since LS is an individual's long-term and cognitive judgment of their living conditions (5,6), it can be expected to be reported at lower levels in the context of persistent risk perception and insecurity.

H1: Fear of experiencing violence in the workplace negatively affects LS.

PR is approached as a process-based concept, defined as the capacity to maintain and regain functionality in the face of adversity/risk; it is described as a context-sensitive and developable adaptation system (11-14). The fact that FFVW is a significant stressor in healthcare services (16,18) and that

resilience in the context of violence is discussed as an explanatory mechanism for outcomes such as depressive symptoms, indicates that there will be a meaningful relationship between the level of fear and the capacity for resilience (15).

H2: There is a significant relationship between fear of workplace violence and PR.

LS is a multi-determinant indicator of subjective well-being shaped by the interaction of internal psychological resources and external conditions (7-10) . Findings indicate that resilience plays a protective/restorative role in maintaining well-being under stressful conditions (38,39), and the nature of resilience is related to adaptation processes (13,14) suggests that higher levels of resilience will be associated with higher LS.

H3: PR positively affects LS.

It has been associated with indicators of distress such as fear, depressive symptoms, anxiety, and impaired sleep quality (17,18) and has been examined in conjunction with burnout processes through the depletion of psychological resources under perceived threat (16). Resilience, on the other hand, can function as a mediating and/or buffering mechanism in stress–health relationships; it can explain individual differences and the level of adverse effects in the context of workplace violence (15). Furthermore, findings indicate that resilience can mediate relationships between variables such as sleep quality and mental well-being (30,31). This framework predicts that the effect of FV on LS involves an indirect pathway, partly mediated by PR.

H4: PR mediates the effect of fear of workplace violence on LS.

Materials and Methods

This cross-sectional study examines the relationship between fear of workplace violence and LS among healthcare workers and tests the mediating role of PR in this relationship. The research was conducted at Balıkesir University Faculty of Medicine Hospital and used convenience sampling due to time and access constraints. Data were collected between March 15, 2024, and April 15, 2025. A total of 173 healthcare workers were included in the study. Data were collected via online and face-to-face surveys. All scales used in the study were rated on a 5-point Likert scale (1=strongly disagree; 5=strongly agree).

Fear of workplace violence was assessed using the scale developed by Rogers and Kelloway (41); the Turkish validity and reliability study of the scale was conducted by Akbolat et al. (42). PR was measured using a 6-item scale developed by Smith et al. (43); the Turkish validity and reliability study was conducted by Doğan (44). LS was assessed using a 5-item scale developed by Diener et al. (5); the Turkish validity and reliability study was conducted by Dağlı and Baysal (45). Scale scores were calculated by averaging the items within each scale, and these average scores were used in the analyses.

To assess potential common method bias in self-report measures, Harman's one-factor test was applied. In this context, a non-rotated principal component analysis was conducted, and the variance explained by the first factor was interpreted with reference to the commonly used 50% threshold in the literature.

Descriptive statistics (means, standard deviations, counts, and percentages) were calculated. The relationships between the variable "experiencing violence" and the socio-demographic variables were evaluated using the chi-square test, and the effect size was reported using Cramer's V. The relationships between continuous variables were examined using Pearson's correlation coefficient. Mediation analysis was performed using Hayes' PROCESS macro (Model 4, v5.0). In this model, fear of workplace violence, PR, and LS were defined as the independent variable (X), the mediator variable (M), and the dependent variable (Y), respectively. The indirect effect was assessed using 5000 bootstrap samples and 95% bootstrap confidence intervals. The variables gender, age range, seniority, educational status, marital status, and exposure to violence were controlled for in the analyses. The level of statistical significance was set at $p<0.05$. The study received ethical approval from the Scientific Ethical Evaluation Board of Selçuk University's Faculty of Economics and Administrative Sciences on (decision no: 03/26, date: 05.03.2024). Written informed consent was obtained from all participants before data collection. Participation was voluntary; participants were informed that they could withdraw from the study at any time without consequence.

Results

Since validity and reliability studies in Turkish had already been conducted for all scales employed in this study, no additional validity analyses were performed. Reliability (Cronbach's alpha) coefficients and descriptive statistics for the scales are presented in Table 1. Accordingly, the Cronbach's alpha coefficients are 0.879 for FV, 0.706 for PR and 0.894 for LS. The mean scale scores were 1.945±0.604 for FV, 3.646±0.482 for PS, and 4.183±0.634 for LS (Table 1).

Table 1. Reliability analysis and means of the scales

Scale	Mean	SD	Cronbach's alpha
Violence fear	1.945	0.604	0.879
Psychological resilience	3.646	0.482	0.706
Life satisfaction	4.183	0.634	0.894

SD: Standard deviation

The socio-demographic characteristics of the participants are presented in Table 2. Of the participants, 68.8% were male (n=119) and 31.2% were female (n=54). Of the participants, 57.8% (n=100) were married and 42.2% (n=73) were single. By seniority, 41.6% of participants were in the 0–5-year group (n=72) and 42.2% were in the 6–10-year group (n=73). Regarding the "experiencing violence" variable, 24.3% of participants answered yes (n=42), and 75.7% answered no (n=131) (Table 2).

The relationship between exposure to violence and socio-demographic variables was examined using chi-square analysis, and the results are presented in Table 3. The analyses indicate a significant association between experience of violence and educational status (χ^2=9.757; df=3; p=0.021). Similarly, a significant relationship was found between the experience of violence and age group (χ^2=8.114; df=3; p=0.044). The relationships between experiencing violence and gender (p=0.167), marital status (p=0.061), and years of employment (p=0.075) were not statistically significant. Effect sizes were reported using Cramér's V, indicating small to moderate effects for age group (V=0.217) and educational status (V=0.237).

The findings of the correlation analysis conducted to determine the relationships between variables are presented in Table 4. Based on these findings, significant relationships were observed: a significant negative relationship between Fear of Experiencing Violence at Work and PR (r=-0.382; p<0.001); a significant negative relationship between Fear of Witnessing Violence (FWV) at work and LS (r=-0.437; p<0.001); and a significant positive relationship between PR and LS (r=0.262; p<0.001) (Table 4).

Regression-based results and path coefficients from the mediation analysis are shown in Table 5. Accordingly, FWV negatively predicts the mediating variable, PR [B=-0.319, standard error (SE)=0.059, t=-5.450; R^2=0.278]. The direct effect of FV on LS is also negative (B=-0.517, SE=0.079, t=-6.520; R^2=0.363). However, the predictive power of PD on LS is not statistically significant when SK and control variables are included in the model (B=0.052, SE=0.099, t=0.531). In the total effect model, the effect of SK on YD continues to be negative (B=-0.534, SE=0.073, t=-7.354; R^2=0.362). In all models, gender, marital status, education, age group, seniority, and exposure to violence were included as covariates.

Discussion

This study demonstrated that fear of workplace violence among healthcare workers is negatively associated with LS; fear is inversely related to PR, while PR is positively associated with LS. This pattern indicates that fear of workplace violence serves as a stressor that undermines subjective well-being by fostering a perception of constant threat among healthcare workers

Table 2. Socio-demographic findings

Socio-demographic characteristics	Category	Number (n)	Percentage (%)
Gender	Male	119	68.8
	Female	54	31.2
Age ranges	30 years old and under	69	39.9
	31-40	77	44.5
	41-50 years old	20	11.6
	51-60 years old	7	4.0
Education	High school	28	16.2
	Associate degree	26	15.0
	Bachelor's degree	62	35.8
	Graduate (master's, doctorate)	57	32.9
Marital status	Married	100	57.8
	Single	73	42.2
Years of work experience	0-5 years	72	41.6
	6-10 years	73	42.2
	11-15 years	9	5.2
	16-20 years	12	6.9
	21 years and over	7	4.0
Experiencing violence	No	131	75.7
	Yes	42	24.3

Table 3. Relationship between experiencing violence and sociodemographic variables					
Variable	**X^2**	**df**	**p**	**Cramer's V**	**n**
Gender	1,908	1	0,167	0,105	173
Age group	8,114	3	0,044	0,217	173
Educational status	9,757	3	0,021	0,237	173
Marital status	3,516	1	0,061	0,143	173
Year of study	8,496	4	0,075	0,222	173

Table 4. Correlations between variables			
Variable 1	**Variable 2**	**r**	**p**
Fear of experiencing violence in the workplace	Psychological resilience	-0.382	<0.001
Fear of experiencing violence at work	Life satisfaction	-0.437	<0.001
Psychological resilience	Life satisfaction	0.262	<0.001

(1,2,17,18). However, because the study is cross-sectional, the findings should be interpreted at the "correlation" level, and claims about directionality and causality should be avoided.

The negative impact of FV on LS is consistent with the view, emphasized in the violence literature, that future uncertainty and threat assessments can determine psychological outcomes, independent of exposure (1). Intensive patient/patient relative contact in healthcare services, rapid decision-making, and conditions of uncertainty may make risk assessments more visible, negatively affecting individuals' general cognitive judgments about their lives (5,6). In this context, when the findings are evaluated alongside reports linking fear to depressive symptoms, anxiety, and sleep disturbance, they suggest that decreased LS may be part of a broader distress profile (17,18). Therefore, assessments of fear levels should be approached not only through "exposure to violence" but also through "perceived risk and control" processes.

Mediation analyses did not support the expected mediating role of PR in the relationship between fear LS. Findings indicate that fear is related to resilience levels; however, the predictive power of resilience for LS is no longer significant when fear and control variables are considered. In other words, the effect of fear on LS largely persists; resilience does not mediate this relationship through a statistically significant indirect effect. This result suggests that, although resilience is generally a valuable resource for well-being, it may not be the primary channel through which fear affects broader outcomes such as LS. This finding should not be interpreted as "resilience is ineffective." Rather, it should be interpreted as in this sample and this model, it has not been supported as an indirect mechanism in the fear-LS pathway.

Possible explanations for this result can be addressed under three headings. First, because LS is a holistic assessment that includes non-work domains, the effect of fear may be mediated by contextual processes such as organizational safety climate, perceived prevention capacity, management support, and perceived fairness rather than by resilience (23).

Second, the role of resilience may vary by outcome; resilience may play a more pronounced mediating/buffering role in psychological outcomes that are "closer" to LS (e.g., anxiety, sleep quality, burnout) (30,31). Third, the presence of numerous covariates in the model may have reduced the specific contribution of resilience by explaining a portion of the variance in LS. This is consistent with the PD coefficient becoming insignificant in the multivariate model (including SK and covariates), despite the PD-LS relationship being significant at the correlation level; that is, the bivariate relationship may weaken at the multivariate level due to shared variance and possible measurement error or short scale length. Therefore, future research should model resilience alongside "organizational context" and "proximal psychological processes" to distinguish the mechanisms at play more clearly. Additionally, the possible "moderator" role of resilience (e.g., buffering the effect of fear on LS) should be tested; moderation may be a more plausible explanation than mediation.

When evaluating the relationships between exposure to violence and socio-demographic variables, analysis revealed significant associations with specific demographic characteristics. This finding provides a framework consistent with the literature, indicating that violence risk and exposure patterns may vary according to demographic/occupational characteristics (20,21). The variation in LS across age groups, meanwhile, aligns with developmental findings indicating that LS can change across life stages (35,37,46). However, as these analyses are cross-

Table 5. Regression-based results in the mediation analysis								
Outcomes	**Predictors**	**R**	**R**	**F (df1, df2)**	**B**	**SE**	**t**	**95% CI**
PD (mediator)	SK	0.528	0.278	F (14,158)=4.353	-0.319	0.059	-5.450	(-0.435, -0.204)
YD (outcome)	SK	0.603	0.363	F (15,157)=5.975	-0.517	0.079	-6.520	(-0.673, -0.360)
YD (outcome)	PD	0.603	0.363	F (15,157)=5.975	0.052	0.099	0.531	(-0.143, 0.247)
YD (total effect)	SK	0.602	0.362	F (14,158)=6.411	-0.534	0.073	-7.354	(-0.677, -0.390)
All models include covariates (gender, marital status, education, age group, tenure, exposure to violence), SE: Standard error, CI: Confidence interval								

sectional and categorical, the findings should be interpreted as a distribution or pattern rather than for determining risk.

From clinical and managerial perspectives, the results suggest that "post-event" approaches alone may not be sufficient to combat violence at the hospital level; processes aimed at reducing uncertainty and perceived risk, which fuel fear, need to be strengthened. The accessibility of violence reporting mechanisms, post-event psychosocial support, management feedback, and the visibility of safety practices may be critical in reducing fear levels (23). In this context, resilience-building programs may be supportive; however, since findings show that fear's direct impact on LS remains strong, multi-component interventions that strengthen organizational prevention and the safety climate appear to be a higher priority (1,16,18). In practice, "individual resource enhancement" approaches should be positioned as complementary to, rather than as replacements for, "organizational risk reduction" strategies.

Study Limitations

These findings should be considered in light of certain limitations. Because the study has a cross-sectional design, the relationships between variables cannot be interpreted causally; longitudinal studies that assess temporal sequence would provide stronger evidence. Second, because the sample was selected from a single center using convenience sampling, the generalizability of the results to other hospitals and units is limited. Third, the data are self-reported; this may increase the risk of social desirability bias and standard method bias. Although the single-factor finding does not indicate a clear bias, validation with multiple measures would be helpful. Fourth, the "experiencing violence" variable could only be partially captured by the model, because the model did not include details such as the source, type, frequency, or timing of incidents. Finally, since organizational context variables that can shape the level of fear (e.g., perceived prevention capacity, safety climate, management support, perceived justice) were not measured, the possible role of these factors in the fear-well-being relationship could not be directly tested. Additionally, assessing PR using a brief scale may underrepresent context-sensitive components (e.g., team support, coping resources); more comprehensive measures may capture different manifestations of mediating or buffering effects.

Conclusion

This study shows that fear of workplace violence among healthcare workers is negatively related to LS. The findings reveal that PR decreases as fear levels increase and that LS increases as PR increases. However, mediation analyses indicate that PR is not a statistically significant indirect mechanism in the relationship between fear and LS; the effect of fear on LS is mainly direct. In this context, it seems appropriate to shift the priority of interventions toward processes that reduce the organizational risk perception that fuels fear (e.g., safety climate, perceived prevention capacity, management support).

These findings emphasize that post-event approaches alone may not be sufficient to combat violence and highlight the importance of institutional arrangements that reduce employees' perceptions of future threats. Strengthening violence prevention policies, increasing the visibility of security practices, making reporting mechanisms accessible and reliable, and activating post-event psychosocial support can help maintain well-being indicators. In practice, institutional packages that integrate prevention, reporting, feedback, and psychosocial support can be considered a more rational goal than initiatives focused solely on individual resilience. Individual programs should be positioned as complementary to this effort.

Future research should use multicenter samples, longitudinal designs, and measures that assess the nature of violence exposure (type, frequency, source, and timing) in detail. Furthermore, testing models that include proximal psychological processes (e.g., anxiety, sleep disturbance, burnout) and organizational context variables (e.g., safety climate, management support, perceived fairness), which may mediate the effect of fear on well-being, will contribute to a clearer understanding of the mechanisms. In particular, the mediating or moderating role of safety climate/perceived prevention capacity and the "buffering" effect of PR (fear × resilience) should be tested in subsequent studies using interaction terms.

Ethics

Ethics Committee Approval: The study received ethical approval from the Scientific Ethical Evaluation Board of Selçuk University's Faculty of Economics and Administrative Sciences on (decision no: 03/26, date: 05.03.2024).

Informed Consent: Written informed consent was obtained from all participants before data collection. Participation was voluntary; participants were informed that they could withdraw from the study at any time without consequence.

Footnotes

Authorship Contributions

Concept: M.F., M.Ö., Design: M.F., M.Ö., Data Collection or Processing: M.F., M.Ö., Analysis or Interpretation: M.F., M.Ö., Literature Search: M.F., M.Ö., Writing: M.F., M.Ö.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Barling J, Rogers AG, Kelloway EK. Behind closed doors: in-home workers' experience of sexual harassment and workplace violence. J Occup Health Psychol. 2001;6:255-69.
2. Chataway M, Duff J, Xu G. Introducing a new model of fear of workplace violence: examining nurses and midwives' perceptions of victimization risk in their workplaces. Vict Offender. 2025;21:371-90.
3. Demirbaş H, Sağlam Gürmen E, Yorgancıoğlu M. Healthcare under attack. Eurasian J Emerg Med. 2025;24:183-9.
4. Tekin FC, Selcen Öcal F. Attacks on health care worldwide: 5-year review. Eurasian J Emerg Med. 2023;22:211-5.
5. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess. 1985;49:71-5.
6. Pavot W, Diener E. Review of the satisfaction with life scale. Psychol Assess. 1993;5:164-72.
7. Headey B, Kelley J, Wearing A. Dimensions of mental health: life satisfaction, positive affect, anxiety and depression. Soc Indic Res. 1993;29:63-82.
8. McAdams KK, Lucas RE, Donnellan MB. The role of domain satisfaction in explaining the paradoxical association between life satisfaction and age. Soc Indic Res. 2012;109:295-303.
9. Bhattacharyya KK, Molinari V, Gupta DD, Hueluer G. The role of life satisfaction and optimism for successful aging in mid and late life. J Appl Gerontol. 2025;44:428-38.
10. Szcześniak M, Mazur P, Rodzeń W, Szpunar K. Influence of life satisfaction on self-esteem among young adults: the mediating role of self-presentation. Psychol Res Behav Manag. 2021;14:1473-82.
11. Denckla CA, Cicchetti D, Kubzansky LD, Seedat S, Teicher MH, Williams DR, et al. Psychological resilience: an update on definitions, a critical appraisal, and research recommendations. Eur J Psychotraumatol. 2020;11:1822064.
12. Hartmann S, Backmann J, Newman A, Brykman KM, Pidduck RJ. Psychological resilience of entrepreneurs: a review and agenda for future research. J Small Bus Manag. 2022;60:1041-79.
13. Fletcher D, Sarkar M. Psychological resilience: a review and critique of definitions, concepts, and theory. Eur Psychol. 2013;18:12-23.
14. Vella SL, Pai N. A theoretical review of psychological resilience: defining resilience and resilience research over the decades. Arch Med Health Sci. 2019;7:233.
15. Li C, Fu C. Workplace violence and depressive symptoms: the mediating role of fear of future workplace violence and burnout among Chinese nurses. BMC Psychiatry. 2024;24:379.
16. Fu C, Ren Y, Wang G, Shi X, Cao F. Fear of future workplace violence and its influencing factors among nurses in Shandong, China: a cross-sectional study. BMC Nurs. 2021;20:123.
17. Fu C, Lv X, Cui X, Huang M, Cao F. The association between fear of future workplace violence and depressive symptoms among nurses based on different experiences of workplace violence: a cross-sectional study. BMC Nurs. 2023;22:123.
18. Lin PZ, Wang X, Lin YS, Chen LL, Wu BY. Fear of future workplace violence and sleep quality in Chinese clinical nurses: the mediating role of anxiety and depression. BMC Nurs. 2025;24:1051.
19. Fu C, Wang G, Shi X, Ren Y, Cao F. The association between fear of future workplace violence and burnout among nurses in China: a cross-sectional study. J Affect Disord. 2021;293:29-35.
20. Gash V, Blom N. Workplace violence and fear of violence: an assessment of prevalence across industrial sectors and its mental health effects. Scand J Work Environ Health. 2025;51:370-9.
21. Pacheco E, Cunha M, Duarte J. Violence, aggression and fear in the workplace. Eur Proc Soc Behav Sci. 2016;27-41.
22. Mento C, Silvestri MC, Bruno A, Muscatello MRA, Cedro C, Pandolfo G, et al. Workplace violence against healthcare professionals: a systematic review. Aggress Violent Behav. 2020;51:101381.
23. Mueller S, Tschan F. Consequences of client-initiated workplace violence: the role of fear and perceived prevention. J Occup Health Psychol. 2011;16:217-29.
24. Den Hartigh RJR, Hill Y. Conceptualizing and measuring psychological resilience: what can we learn from physics? New Ideas Psychol. 2022;66:100934.
25. Sisto A, Vicinanza F, Campanozzi LL, Ricci G, Tartaglini D, Tambone V. Towards a transversal definition of psychological resilience: a literature review. Medicina (Kaunas). 2019;55:745.
26. Smeeth D, Beck S, Karam EG, Pluess M. The role of epigenetics in psychological resilience. Lancet Psychiatry. 2021;8:620-9.
27. Bozdağ F, Ergün N. Psychological Resilience of healthcare professionals during COVID-19 pandemic. Psychol Rep. 2021;124:2567-86.
28. IJntema RC, Schaufeli WB, Burger YD. Resilience mechanisms at work: the psychological immunity-psychological elasticity (PI-PE) model of psychological resilience. Curr Psychol. 2023;42:4719-31.
29. Dou D, Shek DTL, Tan L, Zhao L. Family functioning and resilience in children in mainland China: life satisfaction as a mediator. Front Psychol. 2023;14:1175934.
30. Aksakal A, Aksakal E, Kaya F, Kerget B, Afşin DE, Araz Ö. Psychological resilience as a mediator between sleep quality and mental well-being in patients with obstructive sleep apnea syndrome. BMC Psychol. 2025;13:979.
31. Yuan Y, Yang J, Huang W, Hu C, Zhang W, Chen B. Relationships among anxiety, psychological resilience, and physical activity in university students: variable-centred and person-centred perspectives. Front Psycho. 2025;16:1694344.
32. Badri MA, Alkhaili M, Aldhaheri H, Yang G, Albahar M, Alrashdi A. Exploring the reciprocal relationships between happiness and life satisfaction of working adults-evidence from Abu Dhabi. Int J Environ Res Public Health. 2022;19:3575.
33. Shin DC, Johnson DM. Avowed happiness as an overall assessment of the quality of life. Soc Indic Res. 1978;5:475-492.
34. Wodarz R, Rogowska AM. The moderating effect of body appreciation on the relationship between self-esteem and life satisfaction. Eur J Investig Health Psychol Educ. 2024;14:870-87.
35. Daly M. Cross-national and longitudinal evidence for a rapid decline in life satisfaction in adolescence. J Adolesc. 2022;94:422-34.
36. Mpofu E, Zhan RF, Yin C, Brock K. Qualities of older adults' family and friendship relationships and their association with life satisfaction. Geriatrics (Basel). 2024;9:49.
37. Haider SI, Ahmed F, Pasha H, Pasha H, Farheen N, Zahid MT. Life satisfaction, resilience and coping mechanisms among medical students during COVID-19. PLoS One. 2022;17:e0275319.
38. Delhey J, Hess S, Boehnke K, Deutsch F, Eichhorn J, Kühnen U, et al. Life satisfaction during the COVID-19 pandemic: the role of human, economic, social, and psychological capital. J Happiness Stud. 2023;24:2201-22.
39. Chen C, Bian F, Zhu Y. The relationship between social support and academic engagement among university students: the chain mediating effects of life satisfaction and academic motivation. BMC Public Health. 2023;23:2368.
40. Pavot W, Diener E. The satisfaction with life scale and the emerging construct of life satisfaction. J Posit Psychol. 2008;3:137-52.

41. Rogers KA, Kelloway EK. Violence at work: personal and organizational outcomes. J Occup Health Psychol. 1997;2:63-71.
42. Akbolat M, Sezer C, Ünal Ö, Amarat M, Durmuş A. The effects of direct violence and witnessed violence on the future fear of violence and turnover intention: a study of health employees. Curr Psychol. 2021;40:4684-90.
43. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. Int J Behav Med. 2008;15:194-200.
44. Doğan T. Adaptation of the brief resilience scale into Turkish: a validity and reliability study. J Happiness Well-Being. 2015;3:93-102.
45. Dağlı A, Baysal N. Adaptation of The satisfaction with life scale into Turkish: the study of validity and reliability. Electronic Journal of Social Sciences. 2016;15:1250-62.
46. Cho D, Cheon W. Older adults' advance aging and life satisfaction levels: effects of lifestyles and health capabilities. Behav Sci (Basel). 2023;13:293.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.98957
Eurasian J Emerg Med. 2026;25: 246-53

Investigation of the Effect of Laparoscopic Nissen Fundoplication on Sleep Quality: A Pilot Exploratory Study

Şükrü Salih Toprak[1], Ceyhun Pekin[1], Hatice Toprak[2]

[1]Karamanoğlu Mehmetbey University Faculty of Medicine, Department of General Surgery, Karaman, Türkiye
[2]Karamanoğlu Mehmetbey University Faculty of Medicine, Department of Anesthesiology and Reanimation, Karaman, Türkiye

Abstract

Aim: Gastroesophageal reflux disease (GERD) is a significant burden on the healthcare systems of countries. GERD can lead to poor sleep quality, which affects quality of life and worker productivity. The aim of this study was to investigate the effects of laparoscopic Nissen fundoplication (LNF), the surgical treatment for GERD, on sleep quality.

Material and Methods: The study was designed prospectively. The study included 40 patients with extraoesophageal reflux symptoms who underwent LNF. Patients' reflux symptoms and sleep quality were evaluated using the Reflux Symptom index (RSI) and the Pittsburgh Sleep Quality index (PSQI) preoperatively and at 1 month postoperatively.

Results: There was a statistically significant decrease in RSI and PSQI in the postoperative 1st month compared with preoperative values. Preoperative and 1st-month postoperative values for RSI were 23.775 and 4.775, respectively, and for PSQI were 11.250 and 5.580, respectively ($p<0.05$). There was no correlation between decreases in RSI and PSQI.

Conclusion: At 1 month postoperatively, LNF improved extraoesophageal reflux symptoms and sleep quality. No correlation was found between improvements in sleep quality and reflux symptoms.

Keywords: Poor sleep quality, laparoscopic Nissen fundoplication, extraoesophageal reflux symptoms

Introduction

Gastroesophageal reflux disease (GERD) is the most common disease of the gastrointestinal (GI) tract (1). It has an important share in the health expenditures of countries (2). Gastroesophageal reflux (GER) is a clinical condition associated with reflux of gastric contents back into the oesophagus (1), causing uncomfortable symptoms and complications (3) and decreased quality of life (4). Low lower oesophageal sphincter (LES) pressure (usually less than 10 mmHg) and increased intragastric pressure are accepted mechanisms for the development of GER (5).

Laryngopharyngeal reflux (LPR) is the name given to the clinical condition in which GERD has atypical symptoms (hoarseness, sore throat, persistent cough, asthma, laryngospasm and less common symptoms) in the upper aerodigestive system (6). Reflux Symptom index (RSI) was developed by Belafsky et al. (7) to diagnose LPR, and its Turkish reliability and validity were evaluated by Akbulut et al. (8). Laparoscopic Nissen fundoplication (LNF) was introduced in 1991 (9) and attracted interest in the treatment of GERD (10). LNF aims to reduce reflux episodes by anatomically restoring the anti-reflux barrier (11). It provides a significant improvement in LES pressure (12). Studies have found the efficacy of LNF superior to proton pump inhibitors (PPI) for up to 2 years (13,14).

GERD has been associated with sleep disorders and impairments in daily functioning (1,5,15). Sleep disturbance due to GER is an important cause of decreased worker productivity (16). It is thought that there is a bidirectional relationship between GERD and sleep disorders (17). GER has previously been associated with increased sleep disturbance in various studies (15,18,19). Studies showing an

Corresponding Author: Şükrü Salih Toprak MD, Karamanoğlu Mehmetbey University Faculty of Medicine, Department of General Surgery, Karaman, Türkiye
E-mail: drsstbir@gmail.com **ORCID ID:** orcid.org/0000-0002-9695-1841

Cite this article as: Toprak ŞS, Pekin C, Toprak H. Investigation of the effect of laparoscopic Nissen fundoplication on sleep quality: a pilot exploratory study. Eurasian J Emerg Med.2026;25: 246-53.

Received: 13.01.2026
Accepted: 22.03.2026
Published: 30.03.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

increased incidence of GER in patients with sleep disorders are also available in the literature (20,21). Hung et al. (22) found that sleep quality decreased with increased nocturnal acid reflux. Patients with nocturnal reflux symptoms are more likely to experience sleep disturbance (23). Increased nocturnal acid reflux may be considered a preventable cause of poor sleep quality. Previous studies have shown that treatment of nocturnal GER (lifestyle changes, PPIs and antireflux surgery) improves both subjective and objective sleep measures (17). Indications for surgical treatment are failure of medical treatment, side effects of drugs or complications arising from GERD (24). Findings of an increase in quality of life with effective reduction of symptoms in GERD-related LPR were presented in the literature in previous studies (25,26).

The Pittsburgh Sleep Quality index (PSQI) was developed in 1989 for the evaluation of sleep disorders (27). In a previous study, it was reported that 52% of patients diagnosed with GERD experienced sleep disturbance defined as a PSQI score >5.5 (28). Information on the efficacy of LNF for sleep disturbance, one of the extraoesophageal symptoms associated with GERD, is limited in the literature.

In our study, we investigated the efficacy of LNF in improving sleep quality in patients whose extraoesophageal symptoms were evaluated using RSI. We assessed sleep quality using PSQI. The primary aim of our study was to evaluate the change in sleep quality 1 month after LNF. The secondary aim of our study was to examine the relationship between the effects of LNF on RSI and those on PSQI. According to our hypothesis, the change in RSI will parallel the change in PSQI after LNF.

Materials and Methods

This study was designed prospectively. The study was conducted in accordance with the Declaration of Helsinki and the study was approved by Karamanoğlu Mehmetbey University Non-Interventional Clinical Research Ethical Committee (decision no: 08-2022/06, date: 31.08.2022). Permission to conduct the study at Karamanoğlu Mehmetbey University was obtained from the authorised committee of the hospital. The study was registered with the Australian New Zealand Clinical Trials Registry on 02.11.2022 (registration number ACTRN12622001406796). Participants were informed, and informed consent was obtained before participation in the study.

Primary Outcome

Evaluation of the effects of LNF on sleep quality.

Secondary Outcome

The effect of LNF operation on both RSI and PSQI, and the relationship between them, were determined.

Study Population

Between September 2022 and November 2023, the study population consisted of individuals aged 18-65 years with extraoesophageal symptoms who were scheduled for LNF operation for a diagnosis of GER and who had signed written consent documents.

Inclusion

Patients with documented GERD symptoms were included in the study. The diagnosis of GERD was confirmed by the clinical presence of oesophageal and extraoesophageal symptoms, by upper GI endoscopy (to provide evidence of oesophagitis), and by inadequate response to PPI therapy.

Exclusion

Those who had previously undergone anti-reflux surgery and bariatric surgery, those with achalasia and other oesophageal motility disorders, those with chronic diseases that were not in remission (such as uncontrolled diabetes), those who could not read and write Turkish, those with psychiatric disorders, those who did not give consent and those who could not complete the questionnaires were excluded from the study (29).

Evaluation of GER Symptoms

Patients admitted for LNF procedure were evaluated using the RSI in face-to-face interviews on the morning of the operation and at 1 month after discharge, when they presented to our hospital for outpatient follow-up. The purpose of using the RSI was to accurately document the improvement in postoperative extraoesophageal symptoms. The RSI is a 9-item questionnaire that can be completed in less than 5 minutes, with a maximum score of 45 points. The scale for each question ranges from 0 (no problem) to 5 (serious problem). An RSI score higher than 13 indicates the presence of LPR (7,29). The RSI was developed in 2002 (7). Turkish validation was performed in 2020 (8).

Assessment of Sleep Quality

Patients who underwent LNF were evaluated using the PSQI on the morning of the operation and at 1 month after discharge. PSQI was developed in 1989 for the evaluation of sleep disorders (27). In 1996, Turkish validation was performed by Ağargün et al. (30). The PSQI is a 19-item questionnaire that assesses self-reported sleep quality. It evaluates subjective sleep quality, sleep latency (sleep transition time), sleep duration, habitual sleep

efficiency, sleep disorders, sleep medication use and daytime dysfunction in a 1-month period (27).

Surgical Technique

Laparoscopic surgery was performed using a standard technique under general anaesthesia. Induction of anaesthesia was performed using propofol, rocuronium, and fentanyl, dosed according to ideal body weight. Maintenance anaesthesia was provided with 2% sevoflurane in a 50% air-50% O_2 mixture. The intraoperative process was monitored by the anaesthesiologist involved in the study.

The patient was placed on the operating table in the supine position with the legs suspended. Five trocars were placed: at the umbilicus (10 mm) for the optical port; at the right midclavicular line (10 mm) for the liver retractor; at the subxiphoid (10 mm) and the left midclavicular line (5 mm) for the working ports; and at the left anterior axillary line (5 mm) for the assistant port. After excluding the liver, the stomach was pulled from the fundus, and the distal oesophagus was exposed with an ultrasonic dissector. The right and left diaphragmatic crura were exposed while preserving the vagus nerve. The oesophagus was suspended. The right and left crura were approximated with 3 number-0 silk sutures. Then the fundus of the stomach was mobilised by separating it from the spleen through division of the short gastric vessels. The fundus of the stomach was then pulled behind the abdominal oesophagus to create a 360-degree fundoplication. Three 2/0 silk sutures were used for fundoplication. Looseness of the fundoplication was checked, bleeding was controlled, and the operation was completed. All trocars were withdrawn under endocamera guidance (31).

Patients were hospitalised for at least 1 day. Swallowing functions were evaluated before discharge. They were discharged when no problems were observed. A liquid diet was given for the first 10 postoperative days. Solid foods were introduced slowly to prevent dysphagia caused by mucosal oedema (31).

The need for re-hospitalisation, emergency surgery, or elective surgery was assessed at the 1st postoperative month.

Statistical Analysis

The data obtained in the study were analysed using SPSS (Statistical Package for Social Sciences) for Windows, version 22.0. Number, percentage, mean, and standard deviation were used as descriptive statistical methods in the evaluation of the data. A dependent-groups t-test was used to compare repeated measurements. Pearson's correlation analysis was performed between continuous variables.

Results

Forty-four patients who underwent LNF were enrolled. 1 patient was excluded because he was over 65 years of age, and 3 patients did not attend the planned outpatient clinic visit at the end of the 1st postoperative month. Data from 40 patients were analysed. The findings on the patients' descriptive characteristics are presented below.

When the descriptive characteristics of the patients were analysed according to Table 1, the mean age of the sample was calculated as 41.57±10.22 years. When body mass index (BMI) categories were analysed, 72.5% of the patients were overweight, 20% were normal weight and 7.5% were Obesity class 1 (32).

Table 1. Distribution of patients according to identifier characteristics

Groups	Variable (N)	Percentage (%)
Sex		
Male	10	25.0
Female	30	75.0
BMI (kg/m^2)		
18.5<BMI<24.9 (normal weight)	8	20.0
25<BMI<29.9 (overweight)	29	72.5
30<BMI <34.9 (obesity class 1)	3	7.5
Smoking status		
Yes	12	30.0
None	28	70.0
Alcohol use status		
Yes	3	7.5
None	37	92.5

Table 1. Continued		
Groups	**Variable (N)**	**Percentage (%)**
PPI drug use status		
Yes	30	75.0
None	10	25.0
	Mean	**SD**
Age	41,570	10,223
Height (cm)	164,880	6,568
Weight (kg)	72,780	8,707
BMI (kg/m²),	26,737	2,431
BMI: Body mass index (kg/m²), PPI: Proton pump inhibitor, SD: Standard deviation, [1]n (%); Mean ± SD; median (25%-75%)		

This distribution shows that females and overweight individuals predominated in the sample.

The distribution by smoking status showed that 30% of the patients were smokers and 70% were non-smokers. The prevalence of alcohol use was: only 7.5% of patients used alcohol. Analysis of PPI use showed that 75% of the patients used these drugs.

According to the sleep quality components in Table 2, significant improvements were observed postoperatively compared with preoperatively. The overall sleep quality score decreased from a preoperative mean of 11.25 to a postoperative mean of 5.58, and the difference was statistically significant (t=10.725, p=0.000).

Analysis of the subcomponents showed that the subjective sleep quality score decreased from 1.75 preoperatively to 0.50 postoperatively (t=10.648, p=0.000). Postoperative sleep latency, the time to fall asleep, decreased significantly from 2.30 to 1.30 (t=6.774, p=0.000).

Sleep duration also increased postoperatively, and the scores decreased from 1.23 preoperatively to 0.7 postoperatively. This difference was statistically significant (t=3.365, p=0.002). Although a small change was observed in the sleep efficiency component (from 0.88 to 0.60), this difference was not statistically significant (t=1.638, p=0.109). This result may indicate that patients' habits of going to bed and getting up were preserved postoperatively.

The s leep disturbance score decreased from 2.90 preoperatively to 1.73 postoperatively, and this difference was statistically significant (t=5.262, p=0.000).

A postoperative decrease in the use of sleeping pills (0.65 before, 0.33 after) was observed, and this decrease was statistically significant (t=2.481, p=0.018).

A postoperative decrease was observed in the use of sleeping pills (from 0.65 to 0.33), and this decrease was statistically significant (t=2.481, p=0.018).

Finally, a significant improvement in daytime dysfunction was observed (t=6.260, p=0.000). These data may support the conclusion that the operation improved patients' overall sleep quality.

Results of the reflux assessment (Table 3) show that the operation resulted in a significant improvement in reflux symptoms. T he mean reflux assessment score was 23.775 before the operation and decreased to 4.775 after the operation. This difference was statistically significant (t=15,149, p=0,000).

According to the correlation analysis in Table 4 and Figure 1, a strong and positive relationship was found between the before and after measures of sleep quality (r=0.713, p=0.000). This correlation suggests that individuals show continuity in sleep quality. Similarly, the before-and-after measures of reflux assessment show a significant correlation (r=0.434, p=0.005). There is a moderate positive correlation between previous measurements of sleep quality and reflux assessment (r=0.511, p=0.001). A significant correlation was also found between the subsequent sleep quality and reflux assessment measurements (r=0.579, p=0.000), suggesting that changes in sleep quality may be related to reflux symptoms. Although there was no significant correlation between change in sleep quality and change in reflux assessment (r=0.161, p=0.320), the change in reflux assessment was strongly and negatively correlated with the previous reflux assessment score (r=0.807, p=0.000), suggesting that changes in reflux symptoms may differ depending on the initial reflux level.

Complications were evaluated in individuals who underwent LNF for a diagnosis of GERD. There was no need for re-hospitalisation, emergency surgery, or elective surgery in the 1st postoperative month.

Table 2. Sleep quality components

	Preoperative		Postoperative 1st month		N	t	p
	Mean	SD	Mean	SD			
PSQI score	11,250	3,835	5,580	4,722	40	10,725	0,000*
Subjective sleep quality	1,750	0,776	0,500	0,784	40	10,648	0,000*
Sleep latency	2,300	1,018	1,300	1,181	40	6,774	0,000*
Sleep duration	1,230	1,074	0,700	0,966	40	3,365	0,002*
Habitual sleep activity	0,880	1,202	0,600	1,033	40	1,638	0,109
Sleep disorders	2,900	0,496	1,730	1,450	40	5,262	0,000*
Sleep medication use	0,650	1,145	0,330	0,764	40	2,481	0,018*
Daytime dysfunction	1,550	1,300	0,420	0,813	40	6,260	0,000*

PSQI: Pittsburgh Sleep Quality index, SD: Standard deviation, [1]n (%); Mean ± SD; median (25%-75%), *The p-value is <0.05

Table 3. Reflux evaluation measurements

	Preoperative		Postoperative 1st month		N	t	p
	Mean	SD	Mean	SD			
RSI score	23,775	8,657	4,775	5,196	40	15,149	0,000*

RSI: Reflux Symptom index, SD: Standard deviation, [1]n (%); Mean ± SD; median (25%-75%), *p-value is below the threshold of <0.05

Table 4. Correlation analysis for preoperative and postoperative 1st month RSI and PSQI values

		Preoperative PSQI score	Preoperative 1st mouth PSQI score	Preoperative RSI score	Postoperative 1st mouth RSI score	Sleep quality change	Reflux evaluation change
Preoperative PSQI score	r	1,000					
	p	0,000					
Preoperative 1 st mouth PSQI score	r	0,713*	1,000				
	p	0,000	0,000				
Preoperative RSI score	r	0,511*	0,494*	1,000			
	p	0,001	0,001	0,000			
Postoperative 1st mouth RSI score	r	0,336*	0,579*	0,434*	1,000		
	p	0,034	0,000	0,005	0,000		
PSQI change	r	-0,140	0,595*	0,111	0,432*	1,000	
	p	0,388	0,000	0,494	0,005	0,000	
RSI change	r	-0,338*	-0,160	-0,807*	0,182	0,161	1,000
	p	0,033	0,324	0,000	0,262	0,320	0,000

RSI: Reflux Symptom index, PSQI: Pittsburgh Sleep Quality index, *p-value is below the threshold of <0.05

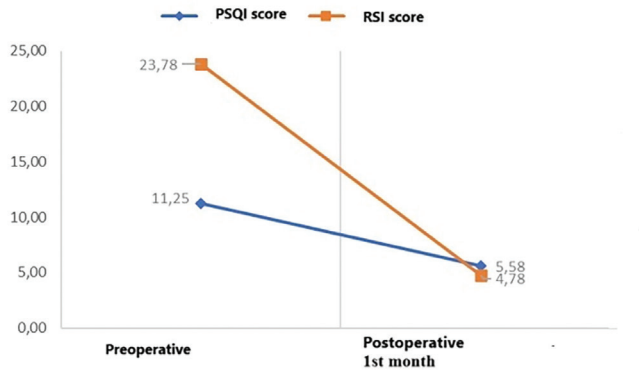


Figure 1. The effect of LNF on RSI and PSQI values preoperatively and at 1 month postoperatively

RSI: Reflux Symptom index, PSQI: Pittsburgh Sleep Quality index, LNF: Laparoscopic Nissen fundoplication

Discussion

In our study, individuals diagnosed with GERD who underwent LNF showed improvement in RSI and PSQI values at the end of the 1st postoperative month compared with preoperative values. No statistically significant correlation was observed between preoperative and postoperative 1st-month changes in RSI and PSQI. This may indicate that LNF has positive but limited effects on improving poor sleep quality. According to the results of our study, habitual sleep efficiency did not show a statistically significant improvement. This result may support the idea that individuals' sleep quality is significantly affected by habits acquired over time.

Obesity, alcohol and smoking are known risk factors for GERD (33). In previous studies, it was found that increased oesophageal acid load caused an increase in symptoms in individuals with obesity (34). The same study concluded that obesity was not effective on oesophageal motor functions (34). When the demographic data of the participants with GERD in our study were analysed, 72.5% of the patients who underwent LNF operation were found to be 25<BMI<29.9 (overweight) (35). The rates of smoking and alcohol use were 30% and 7.5%, respectively. The high prevalence of elevated BMI is consistent with previous studies (36,37). In our study, the rate of alcohol and cigarette use, which we thought was related to the high number of female participants, was low.

In our study, the efficacy of LNF for GERD was evaluated using RSI. RSI is an easily applicable and reproducible tool that can determine the severity of symptoms for LPR in patients diagnosed with GERD (7). The mean RSI values at the end of the preoperative period and the postoperative 1st month were 23.775 and 4.775, respectively (Table 3). This finding suggests that LNF is effective in the treatment of LPR caused by GERD.

In our study, the effects of LNF on sleep quality were evaluated using the PSQI. PSQI scores at the preoperative and postoperative 1st-month time points were compared. The mean total sleep quality scores at the two time points were 11.250 and 5.580 for the preoperative and postoperative 1st month, respectively. This difference was statistically significant (p=0.000). The PSQI is the most commonly used assessment tool among clinical and non-clinical populations (38). It can assess both quantitative aspects of sleep (i.e. sleep duration, sleep latency, number of awakenings) and subjective aspects (depth or restfulness) (38). The total score on the PSQI is between 0 and 21. Scores of 0-5 are associated with good sleep quality and scores of 5 and above are associated with poor sleep quality (27). Our results show that poor preoperative sleep quality improved after LNF. However, the postoperative 1st-month PSQI mean values were above 5. This result may be related with many conditions affecting sleep quality such as obesity, hormonal changes, personal obligations, and differences in emotional state (39). In our study, 80% of our participants had high BMI. The fact that the PSQI values of our participants did not fall below 5 may support the presence of obesity-related sleep disorders.

In our study, 75% of the participants were using PPIs. Patients who underwent upper GI endoscopy and surgery continued their current PPI use. There is no consensus regarding whether PPI use should be discontinued before anti-reflux surgery. In our study, PPI use was not stopped preoperatively considering the risk of development of acid hyper secretion with abrupt discontinuation of PPI (40). We emphasise the importance of evaluating the effects of LNF on sleep quality in patients not using PPI in future studies. Our sample size was insufficient to assess this parameter.

Dysphagia after LNF is an undesirable but possible complication (12). Dysphagia, which is a parameter evaluated by the RSI, may have affected our results, given our short follow-up period. Although this effect could not be ruled out, we observed a statistically significant decrease in RSI after LNF. We emphasise that future studies should include a longer follow-up period to reduce this effect.

Study Limitations

Our study has limitations. Our study was conducted at a single centre, and the operations were performed by a single surgeon. This is advantageous for data standardisation. However, it is disadvantageous for generalising the results. The short follow-up period can be considered another limitation. We preferred to keep our follow-up period short to ensure participants' continued participation in the study and to observe the effect of the acquired sleep habits. We did not calculate the sample size because the number of participants was relatively small. We

hope that professional organisations will develop collaborative approaches to facilitate data collection for specialised surgeries such as LNF.

We hope that future studies will employ methodologies to eliminate the effect of PPI use. We also emphasise that the efficacy of LNF should be evaluated with longer follow-up periods.

Conclusion

This study was conducted to evaluate the effect of LNF on sleep quality. The study found that sleep quality improved in individuals with extraoesophageal symptoms of GERD who underwent LNF. However, this improvement was not as significant as that observed in the extraoesophageal symptoms of GER. Our findings support the notion that sleep quality may be affected by multiple conditions. In particular, we reported that habitual sleep patterns did not change 1 month after LNF. The study supports the idea that sleep quality, which plays an important role in increasing employee productivity, may improve with LNF. We suggest that future studies with longer follow-up periods will contribute to the literature.

This study showed that LNF is an effective treatment protocol for reducing the extraoesophageal symptoms of GERD and contributes to improving sleep quality, although the improvement in sleep quality is smaller than the improvement in extraoesophageal reflux symptoms.

Ethics

Ethics Committee Approval: The study was conducted in accordance with the Declaration of Helsinki and the study was approved by Karamanoğlu Mehmetbey University Non-Interventional Clinical Research Ethical Committee (decision no: 08-2022/06, date: 31.08.2022).

Informed Consent: Participants were informed, and informed consent was obtained before participation in the study.

Acknowledgement: Fatih Sontay contributed to the statistical evaluation of the study. A deep-learning artificial intelligence application was used for English-language editing of the article.

Footnotes

Authorship Contributions

Surgical and Medical Practices: Ş.S.T., C.P., H.P., Concept: Ş.S.T., H.P., Design: Ş.S.T., C.P., H.P., Data Collection or Processing: H.P., Analysis or Interpretation: Ş.S.T., H.P., Literature Search: H.P., Writing: Ş.S.T., H.P.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Fujiwara Y, Arakawa T, Fass R. Gastroesophageal reflux disease and sleep disturbances. J Gastroenterol. 2012;47:760-9. Epub 2012 May 17.
2. Shaheen NJ, Hansen RA, Morgan DR, Gangarosa LM, Ringel Y, Thiny MT, et al. The burden of gastrointestinal and liver diseases, 2006. Am J Gastroenterol. 2006;101:2128-38. Epub 2006 Jul 18.
3. Vakil N, van Zanten SV, Kahrilas P, Dent J, Jones R; Global consensus group. The montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. Am J Gastroenterol. 2006;101:1900-20; quiz 1943.
4. Revicki DA, Wood M, Wiklund I, Crawley J. Reliability and validity of the gastrointestinal symptom rating scale in patients with gastroesophageal reflux disease. Qual Life Res. 1998;7:75-83.
5. Orr WC. Review article: sleep-related gastro-oesophageal reflux as a distinct clinical entity. Aliment Pharmacol Ther. 2010;31:47-56.
6. Hawkins BL. Laryngopharyngeal reflux: a modern day "great masquerader". J Ky Med Assoc. 1997;95:379-85.
7. Belafsky PC, Postma GN, Koufman JA. Validity and reliability of the reflux symptom index (RSI). J Voice. 2002;16:274-7.
8. Akbulut S, Aydinli FE, Kuşçu O, Özcebe E, Yilmaz T, Rosen CA, Gartner-Schmidt J. Reliability and validity of the Turkish reflux symptom index. J Voice. 2020;34:965.e23-8. Epub 2019 Jun 25.
9. Dallemagne B, Weerts JM, Jehaes C, Markiewicz S, Lombard R. Laparoscopic nissen fundoplication: preliminary report. Surg Laparosc Endosc. 1991;1:138-43.
10. Anvari M, Allen C, Marshall J, Armstrong D, Goeree R, Ungar W, et al. A randomized controlled trial of laparoscopic nissen fundoplication versus proton pump inhibitors for the treatment of patients with chronic gastroesophageal reflux disease (GERD): 3-year outcomes. Surg Endosc. 2011;25:2547-54.
11. Yadlapati R, Hungness ES, Pandolfino JE. Complications of antireflux surgery. Am J Gastroenterol. 2018;113:1137-47. Epub 2018 Jun 14.
12. Håkanson BS, Lundell L, Bylund A, Thorell A. Comparison of laparoscopic 270° posterior partial fundoplication vs total fundoplication for the treatment of gastroesophageal reflux disease: a randomized clinical trial. JAMA Surg. 2019;154:479-86.
13. Mahon D, Rhodes M, Decadt B, Hindmarsh A, Lowndes R, Beckingham I, et al. Randomized clinical trial of laparoscopic nissen fundoplication compared with proton-pump inhibitors for treatment of chronic gastro-oesophageal reflux. Br J Surg. 2005;92:695-9.
14. Spechler SJ. Comparison of medical and surgical therapy for complicated gastroesophageal reflux disease in veterans. The Department of Veterans Affairs Gastroesophageal Reflux Disease Study Group. N Engl J Med. 1992;326:786-92.
15. Shaker R, Castell DO, Schoenfeld PS, Spechler SJ. Nighttime heartburn is an under-appreciated clinical problem that impacts sleep and daytime function: the results of a gallup survey conducted on behalf of the American Gastroenterological Association. Am J Gastroenterol. 2003;98:1487-93.
16. Wahlqvist P, Reilly MC, Barkun A. Systematic review: the impact of gastro-oesophageal reflux disease on work productivity. Aliment Pharmacol Ther. 2006;24:259-72.
17. Shibli F, Skeans J, Yamasaki T, Fass R. Nocturnal gastroesophageal reflux disease (GERD) and sleep: an important relationship that is commonly overlooked. J Clin Gastroenterol. 2020;54:663-74.

18. Chen MJ, Wu MS, Lin JT, Chang KY, Chiu HM, Liao WC, et al. Gastroesophageal reflux disease and sleep quality in a Chinese population. J Formos Med Assoc. 2009;108:53-60.
19. Gurges P, Murray BJ, Boulos MI. Relationship between gastroesophageal reflux disease and objective sleep quality. J Clin Sleep Med. 2022;18:2731-8.
20. Liu Y, Wu J, Xiao F, Gu X, Ji L. Correlation and influencing factors between laryngopharyngeal reflux disease and sleep status in patients. Front Surg. 2022;9:845653.
21. Pandey A, Kar SK. Rapid Eye Movement sleep deprivation of rat generates ROS in the hepatocytes and makes them more susceptible to oxidative stress. Sleep Sci. 2018;11:245-53.
22. Hung JS, Lei WY, Yi CH, Liu TT, Chen CL. Association between nocturnal acid reflux and sleep disturbance in patients with gastroesophageal reflux disease. Am J Med Sci. 2016;352:141-5. Epub 2016 May 25.
23. Mody R, Bolge SC, Kannan H, Fass R. Effects of gastroesophageal reflux disease on sleep and outcomes. Clin Gastroenterol Hepatol. 2009;7:953-9. Epub 2009 Apr 16.
24. Slater BJ, Dirks RC, McKinley SK, Ansari MT, Kohn GP, Thosani N, et al. SAGES guidelines for the surgical treatment of gastroesophageal reflux (GERD). Surg Endosc. 2021;35:4903-17. Epub 2021 Jul 19.
25. Catania RA, Kavic SM, Roth JS, Lee TH, Meyer T, Fantry GT, et al. Laparoscopic nissen fundoplication effectively relieves symptoms in patients with laryngopharyngeal reflux. J Gastrointest Surg. 2007;11:1579-87; discussion 1587-8. Epub 2007 Oct 12.
26. Iqbal M, Batch AJ, Moorthy K, Cooper BT, Spychal RT. Outcome of surgical fundoplication for extra-oesophageal symptoms of reflux. Surg Endosc. 2009;23:557-61. Epub 2008 Mar 26.
27. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28:193-213.
28. Fujiwara Y, Kohata Y, Kaji M, Nebiki H, Yamasaki T, Sasaki E, et al. Sleep dysfunction in Japanese patients with gastroesophageal reflux disease: prevalence, risk factors, and efficacy of rabeprazole. Digestion. 2010;81:135-41. Epub 2010 Jan 14.
29. Paranyak M, Patel R. A prospective randomized trial on laparoscopic total vs partial fundoplication in patients with atypical symptoms of gastroesophageal reflux disease. Langenbecks Arch Surg. 2023;408:269.
30. Ağargün MY, Kara H, Anlar Ö. The validity and reliability of the Pittsburgh Sleep Quality index. Turk Psikiyatri Derg. 1996;7:107-15.
31. Nikolic M, Schwameis K, Semmler G, Asari R, Semmler L, Steindl A, et al. Persistent dysphagia is a rare problem after laparoscopic nissen fundoplication. Surg Endosc. 2019;33:1196-205. Epub 2018 Aug 31.
32. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet. 2004;363:157-63.
33. Cheng Y, Kou F, Liu J, Dai Y, Li X, Li J. Systematic assessment of environmental factors for gastroesophageal reflux disease: an umbrella review of systematic reviews and meta-analyses. Dig Liver Dis. 2021;53:566-73. Epub 2020 Dec 16.
34. Rogers BD, Patel A, Wang D, Sayuk GS, Gyawali CP. Higher esophageal symptom burden in obese subjects results from increased esophageal acid exposure and not from dysmotility. Clin Gastroenterol Hepatol. 2020;18:1719-26. Epub 2019 Aug 20.
35. Purnell JQ. Definitions, Classification, and epidemiology of obesity. 2023 May 4. In: Feingold KR, Adler RA, Ahmed SF, Anawalt B, Blackman MR, Chrousos G, editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000.
36. Cai N, Ji GZ, Fan ZN, Wu YF, Zhang FM, Zhao ZF, et al. Association between body mass index and erosive esophagitis: a meta-analysis. World J Gastroenterol. 2012;18:2545-53.
37. Hampel H, Abraham NS, El-Serag HB. Meta-analysis: obesity and the risk for gastroesophageal reflux disease and its complications. Ann Intern Med. 2005;143:199-211.
38. Mollayeva T, Thurairajah P, Burton K, Mollayeva S, Shapiro CM, Colantonio A. The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: a systematic review and meta-analysis. Sleep Med Rev. 2016;25:52-73. Epub 2015 Feb 17.
39. Palacios S, Chedraui P, Sánchez-Borrego R, Coronado P, Nappi RE. Obesity and menopause. Gynecol Endocrinol. 2024;40:2312885. Epub 2024 Feb 11.
40. Katz PO, Dunbar KB, Schnoll-Sussman FH, Greer KB, Yadlapati R, Spechler SJ. ACG Clinical guideline for the diagnosis and management of gastroesophageal reflux disease. Am J Gastroenterol. 2022;117:27-56.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.17999
Eurasian J Emerg Med. 2026;25: 254-60

The Value of Bedside Tests in Differentiating Peripheral and Central Vertigo in the ED

Tansu Akpınar, Şerife Özdinç

Afyonkarahisar Health Sciences University Faculty of Medicine, Department of Emergency Medicine, Afyonkarahisar, Türkiye

Abstract

Aim: This study aims to differentiate peripheral from central vertigo in patients presenting to the emergency department (ED) with complaints of dizziness, using medical history, physical examination, bedside tests, and blood tests.

Materials and Methods: This prospective, observational, cross-sectional study was conducted in the Adult ED at the Afyonkarahisar Health Sciences University Faculty of Medicine between June 24, 2024, and April 24, 2025. Patients aged 18 years and older who presented to the ED with complaints of dizziness and who consented to participate in the study were included. Data on patient demographics, presenting complaints, medical history, bedside test results and scores, neurological examination findings, vital signs, and laboratory results were recorded on a form. Descriptive statistics were used in addition to Chi-square tests, t-tests, and Mann-Whitney U tests for group comparisons. Values with a p-value below 0.05 were considered statistically significant.

Results: Peripheral vertigo was present in 77.2% (n=78) of the 101 patients included in the study. The mean age of patients with central vertigo was 69.79±9.89 years, while the mean age of patients with peripheral vertigo was 52.84±17.07 years (p<0.001). Central vertigo was significantly more prevalent among patients with a history of brain tumors (p=0.033), diabetes mellitus (DM) (p=0.001), and atrial fibrillation (AF) (p=0.016). Conversely, peripheral vertigo was significantly more prevalent among patients with a history of alcohol use (p=0.036). We found that in patients with central vertigo, the STANDING algorithm showed central causes at a significant rate; the head impulse test was negative, the skew test was positive, and vertical nystagmus accompanied it (p<0.001). Laboratory tests showed that levels of glucose (p=0.010), C-reactive protein (CRP) (p=0.002), and the glucose-to-K^+ ratio (p=0.002) were significantly higher in patients with central vertigo. Albumin levels were lower in patients with central vertigo than in those with peripheral vertigo (p<0.001).

Conclusion: The study found that advanced age, acute cerebrovascular accident, brain tumor, DM, AF, history of alcohol use, neurological examination findings, $ABCD^2$, TriAGe+ score, head impulse, nystagmus, test of skew test, and STANDING algorithm, blood glucose, CRP, albumin level, and glucose/K+ ratio as parameters that can be used to differentiate between central and peripheral vertigo.

Keywords: Peripheral vertigo, central vertigo, HINTS test, $ABCD^2$ score, TriAGe+ score, STANDING algorithm

Introduction

Vertigo, defined as dizziness, imbalance, and a spinning sensation, affects 20% of the population and is one of the most common causes (4%) of emergency department (ED) visits (1-3). Vertigo can be caused by many conditions, such as hypoglycemia, hypovolemia, metabolic disorders, cardiac conditions, and ear disorders. While peripheral vertigo causes are identified in 80% of cases, central events such as cerebrovascular accident (CVA), transient ischemic attack (TIA), intracranial hemorrhage (ICH), and malignancy have been detected in 15-20% of cases (4,5). A detailed medical history, physical and neurological examinations, and, if necessary, blood tests and imaging methods are crucial for distinguishing between peripheral and central vertigo. In addition, bedside tests and scoring systems that are not commonly used in ED practice but are cost-free and can distinguish central vertigo, such as the head impulse, nystagmus, test of skew (HINTS) test, STANDING algorithm, $ABCD^2$, and TriAGe+ score, are also available.

Corresponding Author: Tansu Akpınar MD, Afyonkarahisar Health Sciences University Faculty of Medicine, Department of Emergency Medicine, Afyonkarahisar, Türkiye
E-mail: tansucimenoglu@hotmail.com **ORCID ID:** orcid.org/0009-0009-6452-2776

Cite this article as: Akpınar T, Özdinç Ş. The value of bedside tests in differentiating peripheral and central vertigo in the ED. Eurasian J Emerg Med. 2026;25: 254-60.

Received: 24.09.2025
Accepted: 24.03.2026
Published: 30.03.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

The HINTS test is highly specific (82-100%); it is performed by combining the head impulse test (HIT), which evaluates the vestibulo-ocular reflex and corrective eye movements observed in peripheral pathologies; the nystagmus assessment, which evaluates nystagmus that can be reversible, vertical, or rotatory, in central pathologies; and the skew test, which detects a refocusing movement when one eye is closed and then opened, occurring in lesions of the supranuclear nuclei associated with vision in the brainstem (6). The significance of any one of the three is sufficient for central pathologies. A centrally significant test has been shown to be more sensitive than brain diffusion magnetic resonance (BDMR) in diagnosing acute stroke within the first 48 hours of symptoms. (7). The STANDING algorithm is a four-step procedure with high specificity (71.8%-94.3%) and sensitivity (93.4%-100%) for identifying central pathologies. It evaluates spontaneous versus positional nystagmus, nystagmus direction, HIT, and balance (8). The ABCD² score is used to determine stroke risk after a TIA. It considers age, blood pressure, clinical findings, symptom duration, and diabetes status. A score of 0-3 indicates low risk, 4-5 indicates moderate risk, and 6-7 indicates high risk (9). The TriAGe+ score has a similar diagnostic value to that of imaging tests and can easily detect CVA-related vertigo early. It consists of components such as neurological findings, symptom duration, vascular risk factors, age, and vital signs (10). For patients with vertigo, TriAGe+ scores of less than 5 indicate low risk, scores of 5 to 7 indicate moderate risk, scores of 8 to 9 indicate high risk, and scores greater than 9 indicate very high risk.

The study evaluates diagnostic parameters, particularly bedside tests, that differentiate peripheral from central vertigo in patients presenting to the ED with dizziness.

Materials and Methods

This prospective, observational, cross-sectional study was conducted at the Afyonkarahisar Health Sciences University Faculty of Medicine between 06.07.2024 and 04.24.2025 after obtaining approval from the Afyonkarahisar Health Sciences University Faculty of Medicine Clinical Research Ethics Committee (approval no. 2024/4). Our study was conducted prospectively in an adult ED. Patients aged 18 years and older who presented to the ED with complaints of dizziness and consented to participate in the study were included. Those with hypovolemia, hemorrhage, hypoglycemia, arrhythmia, acute coronary syndrome (ACS), electrolyte imbalance, acid-base imbalance, renal failure, intoxication, altered consciousness, dementia, unstable vital signs, acute trauma, refusal to consent, or those who were pregnant or under 18 years of age were excluded. All patients within a specific time period who were eligible for inclusion in our study were included. The 101 patients included in the study had demographic characteristics such as age and gender, presenting complaints (vertigo, hearing loss, tinnitus, ear pain, ear discharge, headache, nausea, vomiting, visual impairment), medical histories, bedside tests and scores (STANDING algorithm, HINTS test, ABCD² score, TriAGe+ score), neurological examination findings, vital signs (blood pressure, pulse), complete blood count, biochemical tests [hemoglobin, platelet (PLT), neutrophils, monocytes, lymphocytes, lactate, Mg^{2+}, glucose, potassium (K^+), CRP, albumin, and certain ratios (neutrophil/lymphocyte ratio (NLR), PLT/lymphocyte, monocyte/lymphocyte, glucose/K^+, CRP/albumin], computerized tomography (CT) and BDMR examination results in indicated patients were recorded in the data collection form we prepared. Patients with advanced age, diabetes mellitus (DM), hypertension (HT), atrial fibrillation (AF), stroke, and a history of central nervous system disorders; those with physical examination findings such as balance disorders, extremity weakness, speech disorders, and facial paralysis; those without a diagnosis of peripheral vertigo who experienced dizziness for the first time and did not respond to medical treatment; and those with central vertigo indicated by the HINTS test, STANDING algorithm, TriAGE score, and ABCD² score underwent central imaging. Patients in whom pathology was detected on central imaging were diagnosed with central vertigo. A third-year emergency medicine resident, blinded to the study, performed bedside tests. They were administered at the bedside in an adult ED. Factors such as patient consciousness level and mobility could affect the feasibility of the test. Only patients able to perform the tests were included in the study.

Statistical Analysis

The SPSS version 22.0 was used for statistical analysis. The data recorded in the data collection form were transferred to the statistical program. For descriptive statistics, the mean, standard deviation, median and maximum values, and percentages were used. To evaluate the normal distribution of the data, histograms, plots, the coefficient of variation, skewness, kurtosis, and normality tests (Kolmogorov-Smirnov, Shapiro-Wilk) were used. To compare groups, the chi-square test (Pearson's Chi-square, Fisher's exact test if the number of observations per cell is less than five) was used for categorical data; the t-test was used for normally distributed data; and the Mann-Whitney U test was used for non-normally distributed data. Values with a p-value below 0.05 were considered statistically significant. All patients within a specific time period who met the inclusion criteria were included in our study. The number of patients was determined using a power analysis. A total of 101 patients meeting our established study criteria were included. ROC analysis could not be performed because the statistical program we used did not have data from a sufficient number of patients.

Results

Peripheral vertigo was detected in 77.2% (n=78) and central vertigo in 22.8% (n=23) of the 101 patients included in the study; central vertigo was confirmed by imaging methods.

Descriptive statistics and comparisons of patient data for the central and peripheral vertigo groups are presented in Tables 1-3. Table 1 includes only categorical data (known diseases, complaints, examination results, bedside tests) that were present. The mean age of the patients included in the study was determined to be 56.87±17.22. A statistically significant difference was found between central and peripheral vertigo in age (p<0.001, t-test). The mean age of patients with central vertigo (69.79±9.89) was higher than that of patients with peripheral vertigo (52.84±17.07). 53 patients (52.5%) were female, and 48 (47.5%) were male. No significant difference in gender distribution was observed between central and peripheral vertigo (p=0.974, Pearson Chi-square).

Comorbidities, smoking, and alcohol use were evaluated. Central vertigo was more common than peripheral vertigo among patients with a history of cerebrovascular disease; this difference was of borderline statistical significance (p=0.059). Central vertigo was significantly more common than peripheral vertigo in patients with a history of brain tumor (p=0.033), DM (p=0.001), or AF (p=0.016). Peripheral vertigo was significantly more common than central vertigo in patients with a history of alcohol use (p=0.036). No significant differences were found between the two vertigo groups with respect to other diseases (p=0.05). When the complaints accompanying the patients' dizziness were examined, the most common associated symptoms were headache (76.2%) and nausea (71.3%), but no significant difference was found between central and peripheral vertigo for hearing loss, tinnitus, earache, headache, and nausea (p=0.05).

On physical examination, a significant difference was detected between central and peripheral vertigo regarding facial asymmetry, weakness, and loss of balance (p<0.001). These findings were more common in patients with central vertigo. With respect to speech disorder, a borderline significant difference was observed between the two groups (p=0.055). Positive STANDING algorithm results, HIT negativity, and positive Skew test results were found more frequently in patients diagnosed with central vertigo than in those diagnosed with peripheral vertigo (p<0.001). Nystagmus was more frequently detected as vertical in patients with central vertigo than in those with peripheral vertigo (p<0.001). The median ABCD^2 score was 3 (0-7), and the median TriAGe+ score was 5 (0-17). The median ABCD^2 and TriAGe+ scores were significantly higher in patients with central vertigo than in those with peripheral vertigo (p<0.001).

When laboratory results were evaluated, the median values of glucose, glucose/K+ ratio, and CRP were significantly higher in

Table 1. Descriptive statistics of data and comparison of the groups

	All patients mean ± SD, n (%), med (min-max)	**Central vertigo mean ± SD, n (%), med (min-max)**	**Peripheral vertigo mean ± SD, n (%), med (min-max)**	**p**
Age (years)	56.87±17.22	69.79±9.89	52.84±17.07	**<0.001**^{t}
Gender				
Female	53 (52.5%)	12 (52.2%)	41 (52.6%)	0.968^{x2}
Male	48 (47.5%)	11 (47.8%)	37 (%47.4%)	
Resume				
CVA	5 (5%)	4 (16.7%)	1 (1.3%)	0.059^f
Brain tumor	7 (6.9%)	5 (20.8%)	2 (2.6%)	**0.033**^{f}
HT	46 (45.5%)	12 (50%)	34 (44.2%)	0.230^{x2}
HL	17 (16.8%)	4 (16.7%)	13 (16.9%)	0.186^f
DM	37 (36.6%)	14 (58.3%)	23 (29.9%)	**0.00**^{x2}
HF	10 (52.9%)	0 (0%)	10 (13%)	0.116^f
Asthma/COPD	6 (5.9%)	4 (16.7%)	2 (2.6%)	0.102^f
Malignancy	16 (15.8%)	8 (33.3%)	8 (10.4%)	0.314^{x2}
CAD	12 (11.9%)	1 (4.2%)	11 (14.3%)	0.525^f
AF	6 (5.9%)	4 (16.7%)	2 (2.6%)	**0.016**^{f}
Smoke	43 (42.6%)	4 (16.7%)	39 (50.6%)	0.145^f
Alcohol	15 (14.9%)	1 (4.2%)	14 (18.2%)	**0.036**^{f}

^{t}: t-test, ^{x2}: Chi-square test, ^{f}: Fisher's exact test, HT: Hypertension, HL: Hyperlipidemia, DM: Diabetes mellitus, HF: Heart failure, COPD: Chronic obstructive pulmonary disease, CAD: Coronary artery disease, **CVA**: Cerebrovascular accident, AF: Atrial fibrillation, SD: Standard deviation

patients with central vertigo, whereas the median albumin value was lower (p=0.010, p=0.010, p=0.002, p<0.001). No significant differences were found between the central and peripheral vertigo groups with respect to other laboratory results (p>0.05).

Table 2. Descriptive statistics of data and comparison of the groups

	All patients mean ± SD, n (%) med (min-max)	Central vertigo mean ± SD, n (%) med (min-max)	Peripheral vertigo mean ± SD, n (%) med (min-max)	p
Complaint	8 (7.9%)	3 (12.5%)	5 (6.5%)	0.358[f]
Hearing loss	32 (31.7%)	5 (20.8%)	27 (35.1%)	0.478[x2]
Tinnitus	14 (13.9%)	2 (8.3%)	12 (15.6%)	0.595[f]
Earache	77 (76.2%)	17 (70.8%)	60 (77.9%)	0.147[x2]
Headache	72 (71.3%)	19 (79.2%)	53 (68.8%)	0.577[x2]
Nausea	77.28±11.62	79.83±12.47	76.49±11.31	0.166[t]
Examination findings	127.67±19.70	128.12±23.94	127.53±23.94	0.127[t]
Pulse (beats/min)	71.98±14.14	73.12±17.18	71.62±13.16	0.290[t]
SBP (mmHg)	7 (6.9%)	6 (25)	1 (1.3%)	**<0.001[f]**
DBP (mmHg)	8 (7.9%)	8 (33.3%)	0 (0%)	**<0.001[f]**
Facial asymmetry	10 (9.9%)	5 (20.8%)	5 (6.5%)	0.055[f]
Loss of strength	43 (42.6%)	17 (70.8%)	26 (33.8%)	**<0.001[x2]**
Speech impairment				
Loss of balance				

[t]: t-test, [x2]: Chi-square test, [f]: Fisher's exact test, SD: Standard deviation, SBP: Systolic blood pressure

Table 3. Descriptive statistics of data and comparison of the groups

	All patients mean ± SD, n (%) med (min-max)	Central vertigo mean ± SD, n (%). med (min-max)	Peripheral vertigo mean ± SD, n (%) med (min-max)	p
Bedside tests	33 (32.7%)	15 (62.5%)	18 (23.4%)	**<0.001[x2]**
STANDING*	20 (19.8%)	12 (50%)	8 (10.4%)	**<0.001[x2]**
HIT (-)	24 (23.8%)	24 (100%)	0 (0%)	**<0.001[f]**
Nystagmus**	20 (19.8%)	13 (54.2%)	7 (9.1%)	**<0.001[x2]**
Skew test (+)	3 (0-7)	4 (2-7)	2 (0-5)	**<0.001[m]**
$ABCD^2$ score	5 (0-17)	9 (2-17)	4 (0-12)	**<0.001[m]**
TriAGe+ score	16 (7-48)	18 (12-48)	16 (7-48)	0.258[m]
Laboratory	2.05 (1.26-2.49)	1.98 (1.27-2.40)	2.06 (1.26-2.49)	0.444[m]
Lactate (mg/dL)	125 (70-489)	154 (96-489)	118 (70-471)	**0.010[m]**
Mg (mg/dL)	4.60 (0.00-7.00)	6.60 (0.60-381)	4.40 (0.10-188)	**0.002[m]**
Glucose (mg/dL)	3.00 (0.10-3.81)	4.06 (2.53-4.66)	4.42 (2.58-5.64)	**<0.001[m]**
CRP (mg/L)	4.45±0.48	4.41±0.45	4.46±0.49	0.795[t]
Albumin (g/dL)	5.51(1.31-19.99)	5.26 (1.31-19)	5.58 (2.32-17.83)	0.637[m]
K^+ (mg/dL)	1.81±1.01	1.72±1.03	1.83±1.00	0.627[t]
Neutrophils (10^3/uL)	236 (3.24-561)	231 (11-561)	237 (3.24-421)	0.590[m]
Lymphocyte (10^3/uL)	0.60 (0.09-3.83)	0.58 (0.16-1.21)	0.61 (0.09-3.83)	0.369[m]
PLT (10^3/uL)	28.63 (15.56-104)	37.40 (17.78-104)	26.85 (15.56-82.93)	**0.002[m]**
Monocyte (10^3/uL)	1.09 (0.02-126)	1.60 (0.14-126)	1.04 (0.02-45.49)	0.176[m]
Glucose/K^+	3.37 (0.92-29.49)	3.70 (0.92-17.95)	3.32 (0.96-29.49)	0.626[m]
CRP/albumin	143 (2-1196)	144 (2-1196)	143 (5-547)	0.503[m]
Neutrophil/lymphocyte	0.33 (0.08-6.08)	0.32 (0.11-2.33)	0.33 (0.08-6.08)	0.738[m]
PLT/lymphocyte				
Monocyte/lymphocyte				

[x2]: Chi-square test, [t]: t-test, m: Mann Whitney U test, [f]: Fisher's exact test, CVA: Cerebrovascular accident, HT: Hypertension, HL: Hyperlipidemia, DM: Diabetes mellitus, HF: Heart failure, COPD: Chronic obstructive pulmonary disease, CAD: Coronary artery disease, AF: Atrial fibrillation, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, *: Central, **: Vertical, Mg: Magnesium, CRP: C-reactive protein, (mg/dL), K^+: Potassium, PLT: Platelet

CT was performed in 64.6% (n=65) of patients; BDMR imaging was performed in 61.4% (n=62). CT detected ischemia, hemorrhage, or mass in 12 patients, whereas BDMR detected ischemia or mass in 23 patients.

Discussion

Dizziness is one of the most common complaints reported by patients visiting the ED. It accounts for 4% of ED visits in the United States, and more than 30% of people experience dizziness at some point in their lives (1). Vertigo can be considered a sign of cerebral circulation ischemia; patients with acute vertigo who visit the ED have twice the likelihood of experiencing cardiovascular events or stroke compared to those without acute vertigo during a three-year follow-up (11). It has been observed that 2-13.4% of vertigo cases presenting to the ED are due to stroke, and more than 30% of these cases are missed by ED physicians during the initial visit (12). Vertigo can arise from many causes, most of which are benign, leading to diagnostic difficulties and the potential overlooking of central causes. Therefore, a well-taken medical history, a detailed physical examination, bedside diagnostic tests that are noninvasive and cost-free, laboratory tests, and central imaging methods, if necessary, will greatly contribute to distinguishing peripheral from central vertigo, establishing the correct diagnosis, initiating treatment early, and reducing mortality and morbidity due to central causes. In the study, consistent with the literature, 22.8% patients had vertigo due to central causes, and older patients were found to present with central vertigo (13-15). It is known that vertigo is more common in women (13,16). Chronic diseases such as DM, HT, coronary artery disease (CAD), hyperlipidemia (HL), AF, and cancer are risk factors for central vertigo. Among patients with central vertigo, 28% had a history of stroke; there was no significant difference between central and peripheral vertigo in terms of past history of AF, DM, HT, and HL, whereas a history of CVA was significantly associated with central vertigo. Studies have found that a history of HT is associated with a significant difference in central vertigo, but a history of DM, HL, CVA, or AF is not; asthma is associated with Meniere's disease; and asthma and chronic obstructive pulmonary disease are not associated with vertigo (17). In our study, central vertigo was more common in patients with a history of CVA, brain tumor, DM, and AF. These differing results may be related to the population's dietary habits, traditions, socioeconomic status, education, and culture, or to the number of patients included in the study. Similar to our study, studies have found no statistically significant relationship between smoking and central or peripheral vertigo (1,15). However, we found a significantly higher prevalence of peripheral vertigo among patients who consumed alcohol. Another study found a significant relationship between alcohol use and central vertigo (16). These differing results may be due to socioeconomic status, religious beliefs, and the fact that older patients, in whom central vertigo is more common, have more comorbidities and take more medications, which may reduce their alcohol consumption.

Headache and vomiting that accompanied vertigo, as well as auditory symptoms, did not aid in the differential diagnosis. It was determined that headache and vomiting were more common in patients with central vertigo, and there was no significant difference in auditory symptoms (17). When examining the complaints accompanying vertigo in the patients participating in our study, we found that the most common symptoms were headache (76.2%) and nausea (71.3%). Hearing loss and nausea were more common in patients with central vertigo, while tinnitus, earache, and headache were more common in patients with peripheral vertigo. However, we observed no statistically significant difference.

Studies that identify or fail to identify a relationship between high blood pressure and central vertigo when examining pulse and blood pressure, which are the most important components of the physical examination (17,18). A complete neurological examination can provide the clinician with valuable findings for the differential diagnosis of vertigo. Balance disorder, lateral findings, facial asymmetry, speech disorder, ptosis, urinary and fecal incontinence are important physical examination results that support central events (19). In our study, pulse and blood pressure did not help the clinician distinguish vertigo, but neurological examination findings were valuable for identifying central events.

The STANDING algorithm, HINTS test, ABCD^2, and TriAGe+ score, which are bedside tests, are practical, non-invasive, cost-effective, safe, and easy-to-administer methods for differentiating central from peripheral vertigo. Studies have found that HIT, vertical nystagmus, skew test, and consequently the HINTS test resulted in a significant difference in the differentiation of central and peripheral vertigo (20). Our study also found that the STANDING algorithm, HIT negativity, the presence of vertical nystagmus, skew test positivity, and, consequently, the HINTS test could significantly differentiate between patients with central and peripheral vertigo. Similar to our study, studies have found that patients with central pathology scored higher on the ABCD^2 scale (16,20,21). T he median TriAGe+ score was 5 in patients with peripheral vertigo and 9 in patients with central vertigo; the difference between the two groups was significant. A recent study by Toplu and colleagues (22) in an ED also found similar results to ours; the score was 6 for peripheral vertigo and 9 for central vertigo.

Patients presenting to the ED with vertigo typically undergo tests, including a complete blood count, electrolytes, renal and liver function tests, blood glucose, and other blood tests as appropriate to the patient's clinical presentation, to identify potential underlying causes. In our study, we also evaluated whether there was a relationship between central and peripheral vertigo using these tests. Increased lactate levels are directly related to impaired tissue oxygenation, which is associated with ischemic conditions such as cerebellar confusion. Mg is involved in numerous metabolic and biochemical events, such as neuromuscular function, regulation of oxidative stress, glucose, lipid, and protein metabolism, and cell proliferation (23). Mg deficiency is thought to cause many diseases such as migraine, chronic headaches, epilepsy, Alzheimer's disease, Parkinson's disease, and stroke. Glucose, the main energy source of cells, is critical for maintaining cellular metabolism. K^+, the most abundant cation within cells, is important in physiological events such as nerve transmission, heartbeat, muscle contraction, and maintaining normal kidney function. Studies have found that glucose and K^+ may be particularly related to stroke and that these two molecules interact with each other. By utilizing this interaction, the glucose/K^+ ratio was investigated and found to be an early prognostic factor in conditions such as ICH, CVA, and traumatic brain injury (24). CRP, similar to albumin, is an acute phase reactant associated with poor prognosis that also indicates malnutrition in critically ill patients (25). The CRP/albumin ratio may reflect inflammatory conditions and malnutrition (26). CRP may be elevated in vascular diseases and serve as a poor prognostic indicator in ischemic stroke patients. In inflammatory conditions, the neutrophil count increases while the lymphocyte count relatively decreases, and the ratio of these parameters can be used to detect inflammation (27). Furthermore, PLT and monocyte levels are increased in autoimmune diseases and inflammatory conditions. Studies investigating the PLT/lymphocyte ratio are also available. In our study, we evaluated lactate, Mg^{2+}, glucose, CRP, albumin, K^+, neutrophils, lymphocytes, PLT, monocytes, and their ratios to each other. As a result, we found that high glucose and CRP levels, the glucose/K^+ ratio, and low albumin values were significantly different for central vertigo compared to peripheral vertigo. Apart from these findings, we did not observe any significant differences between the two groups in lactate, Mg^{2+}, K^+, neutrophil, lymphocyte, PLT, monocyte, CRP/albumin ratio, NLR, PLT/lymphocyte, and monocyte/lymphocyte measurements included in the study. Studies have found significant differences between central and peripheral vertigo in terms of lymphocyte, neutrophil, NLR, and PLT results, differences in blood glucose, lactate, and CRP levels, no differences in K^+ and PLT results, and significant differences in albumin. We believe that the differing statistical results reported across studies may be related to patients' dietary habits, socioeconomic levels, lifestyles, and existing comorbidities.

Study Limitations

Our study is a single-center, prospective study. Laboratory tests were selected based on the assumption that vertigo could be metabolic in origin. Since imaging studies were not performed on all patients, they (i.e., the imaging studies) were not included in the evaluation. Patients' socioeconomic status, level of awareness, and mobility, as well as the workload and conditions of the ED, all influenced the feasibility of the tests. Since electrocardiograms (ECGs) for all patients were not available, ECGs were not included in the study. ECG was not performed on young patients without comorbidities who had previously experienced vertigo attacks and been diagnosed with peripheral vertigo, had a noisy clinical presentation, and experienced relief after examination and treatment. The diagnosis of ACS was made using ECG, cardiac markers, echocardiography when necessary, and cardiology consultation. Patients diagnosed with ACS were excluded from the study.

Conclusion

Vertigo is one of the most common reasons for visits to the ED and can arise from numerous causes. Early diagnosis and treatment are particularly important in cases of central pathologies that may be life-threatening. According to our study, obtaining a thorough medical history, carefully questioning the patient's background, performing a detailed neurological examination, and primarily conducting noninvasive bedside tests are crucial for diagnosis in patients presenting to the ED with vertigo. In particular, the emergency physician's use of bedside tests can enable the rapid identification of high-risk patients with central vertigo.

Ethics

Ethics Committee Approval: Approval from the Clinical Research Ethics Committee of Afyonkarahisar Health Sciences University Faculty of Medicine (approval no. 2024/4, date: 07.06.2024).

Informed Consent: This prospective, observational, cross-sectional study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: T.A., Ş.Ö., Concept: T.A., Ş.Ö., Design: T.A., Ş.Ö., Data Collection or Processing: T.A., Analysis or Interpretation: T.A., Ş.Ö., Literature Search: T.A., Ş.Ö., Writing: T.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Agrawal Y, Carey JP, Della Santina CC, Schubert MC, Minor LB. Disorders of balance and vestibular function in US adults: data from the National Health and Nutrition Examination Survey, 2001-2004. Arch Intern Med. 2009;169:938-44.
2. Szturm T, Betker AL, Moussavi Z, Desai A, Goodman V. Effects of an interactive computer game exercise regimen on balance impairment in frail community-dwelling older adults: a randomized controlled trial. Phys Ther. 2011;91:1449-62.
3. Newman-Toker DE, Edlow JA. TiTrATE: A novel, evidence-based approach to diagnosing acute dizziness and vertigo. Neurol Clin. 2015;33:577-99, viii.
4. Spiegel R, Kirsch M, Rosin C, Rust H, Baumann T, Sutter R, et al. Dizziness in the emergency department: an update on diagnosis. Swiss Med Wkly. 2017;147:w14565.
5. Herr RD, Zun L, Mathews JJ. A directed approach to the dizzy patient. Ann Emerg Med. 1989;18:664-72.
6. Newman-Toker DE, Kattah JC, Alvernia JE, Wang DZ. Normal head impulse test differentiates acute cerebellar strokes from vestibular neuritis. Neurology. 2008;70:2378-85.
7. Kattah JC, Talkad AV, Wang DZ, Hsieh YH, Newman-Toker DE. HINTS to diagnose stroke in the acute vestibular syndrome: three-step bedside oculomotor examination more sensitive than early MRI diffusion-weighted imaging. Stroke. 2009;40:3504-10.
8. Vanni S, Pecci R, Edlow JA, Nazerian P, Santimone R, Pepe G, et al. Differential diagnosis of vertigo in the emergency department: a prospective validation study of the STANDING Algorithm. Front Neurol. 2017;8:590.
9. Johnston SC, Rothwell PM, Nguyen-Huynh MN, Giles MF, Elkins JS, Bernstein AL, et al. Validation and refinement of scores to predict very early stroke risk after transient ischaemic attack. Lancet. 2007;369:283-92.
10. Kuroda R, Nakada T, Ojima T, Serizawa M, Imai N, Yagi N, et al. The TriAGe+ score for vertigo or dizziness: a diagnostic model for stroke in the emergency department. J Stroke Cerebrovasc Dis. 2017;26:1144-53.
11. Lee SH, Kim JS. Acute diagnosis and management of stroke presenting dizziness or vertigo. Neurol Clin. 2015;33:687-98, xi.
12. Moeller JJ, Kurniawan J, Gubitz GJ, Ross JA, Bhan V. Diagnostic accuracy of neurological problems in the emergency department. Can J Neurol Sci. 2008;35:335-41.
13. Goeldlin M, Gaschen J, Kammer C, Comolli L, Bernasconi CA, Spiegel R, et al. Frequency, aetiology, and impact of vestibular symptoms in the emergency department: a neglected red flag. J Neuro. 2019;266:3076-86.
14. Xing Y, Si L, Zhang W, Wang Y, Li K, Yang X. Etiologic distribution of dizziness/vertigo in a neurological outpatient clinic according to the criteria of the international classification of vestibular disorders: a single-center study. J Neurol. 2024;271:2446-57.
15. Zhu Z, Li Q, Wang L, Xiao C, Wang H, Xu Y. Risk factors and risk model construction of stroke in patients with vertigo in emergency department. Comput Math Methods Med. 2022;2022:2968044.
16. Gerlier C, Hoarau M, Fels A, Vitaux H, Mousset C, Farhat W, et al. Differentiating central from peripheral causes of acute vertigo in an emergency setting with the HINTS, STANDING, and ABCD2 tests: A diagnostic cohort study. Acad Emerg Med. 2021;28:1368-78.
17. Sato T, Kimura A, Yamaguchi H, Honda H, Takahashi T, Harada M, et al. A clinical decision rule to exclude central vertigo in the emergency department: A prospective, multicenter, observational study. GHM Open. 2024;4:11-7.
18. Honda S, Inatomi Y, Yonehara T, Hashimoto Y, Hirano T, Ando Y, et al. Discrimination of acute ischemic stroke from nonischemic vertigo in patients presenting with only imbalance. J Stroke Cerebrovasc Dis. 2014;23888-95.
19. Furman JM, Barton JJS, FRCPC. Evaluation of the patient with vertigo. Updated, 2022.
20. Liu X, Li Z, Ju Y, Zhao X. Application of bedside HINTS, ABCD^2 score and truncal ataxia to differentiate cerebellar-brainstem stroke from vestibular neuritis in the emergency room. Stroke Vasc Neurol. 2024;9:685-92.
21. Kmetonyova S, Paulasova Schwabova J, Sramkova T, Dankova M, Olserova A, Petrzalka M, et al. Posterior circulation stroke diagnosis in unselected group of acutely dizzy patients. Clin Neurol Neurosurg. 2023;224:107541.
22. Toplu ACO, Aslan IK, Akoglu EU, Ozturk TC. The role of the HINTS exam, TriAGe+ score, and ABCD2 score in predicting stroke in acute vertigo patients in the ED. Am J Emerg Med. 2025;91:110-7.
23. Fiorentini D, Cappadone C, Farruggia G, Prata C. Magnesium: biochemistry, nutrition, detection, and social impact of diseases linked to its deficiency. Nutrients. 2021;13:1136.
24. Lu Y, Ma X, Zhou X, Wang Y. The association between serum glucose to potassium ratio on admission and short-term mortality in ischemic stroke patients. Sci Rep. 2022;12:8233.
25. Devran O, Karakurt Z, Adıgüzel N, Güngör G, Moçin OY, Balcı MK, et al. C-reactive protein as a predictor of mortality in patients affected with severe sepsis in intensive care unit. Multidiscip Respir Med. 2012;7:47.
26. Carriere I, Dupuy AM, Lacroux A, Cristol JP, Delcourt C; Pathologies Oculaires Liées à l'Age Study Group. Biomarkers of inflammation and malnutrition associated with early death in healthy elderly people. J Am Geriatr Soc. 2008;56:840-6.
27. Kolaczkowska E, Kubes P. Neutrophil recruitment and function in health and inflammation. Nat Rev Immunol. 2013;13:159-75.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Letter to the Editor
DOI: 10.4274/eajem.galenos.2026.95777
Eurasian J Emerg Med. 2026;25: 261-2

Management of Patients with Severe Acidosis in the Emergency Department

Sema Ayten

İstanbul Göztepe Prof. Dr. Süleyman Yalçın City Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye

Keywords: Acidosis, sodium bicarbonate therapy, etiology

Dear Editor,

I read with great interest the article titled "Outcome of Cardiac Arrest and Non-Cardiac Arrest Patients with Severe Acidosis in the Emergency Department: A Retrospective Cohort Study" written by Öztürk et al. (1), published in the latest issue of the Eurasian Journal of Emergency Medicine. This study, which focuses on the management and mortality predictors of patients with severe acidosis in the emergency department, provides very valuable data for our clinical practice. I congratulate the authors for their comprehensive work.

In order to understand the findings of the study more deeply and strengthen its contribution to clinical practice, I would like to draw attention to the following two critical points.

Treatment Heterogeneity and Bicarbonate Use

In the study, the relationship between severe acidosis (pH<7.1) and high mortality was clearly revealed. However, the effect of sodium bicarbonate treatment on outcomes in this patient group remained unclear. In retrospective studies, the decision to administer bicarbonate often depends on the physician's preference, which can create significant heterogeneity in outcomes. Bicarbonate itself is known to increase paradoxical intracellular acidosis, especially in lactic acidosis, leading to side effects such as hypernatremia. The authors' analysis of the difference in mortality between patients receiving bicarbonate and those not receiving bicarbonate would be much more enlightening for readers, especially in light of discussions in the current literature (such as the BICAR-ICU studies) (1-4).

Classification of Acidosis Etiology

Although the prognostic value of lactate levels is emphasized in the article, it is important to note that the etiological origin of acidosis (e.g.; type A lactic acidosis, ketoacidosis, or uremic acidosis) has not been detailed. A subgroup analysis, especially on the anion gap, can help us understand which type of acidosis is more lethal and which group is more susceptible to emergency intervention.

Conditions such as lactic acidosis (type A or type B) and ketoacidosis differ in terms of their underlying pathophysiology and response to treatment. For example, acidosis is generally better tolerated in ketoacidosis, while lactic acidosis (particularly shock-induced type A) is associated with a poorer prognosis (5).

The authors' further classification of acidosis etiology using anion gap calculations and comparison of the mortality rates of these subgroups will add a deeper clinical significance to the findings.

In this case, the authors' inclusion of the limitation that anion gap values may not always be available due to missing records in the retrospective data would further strengthen the article.

In conclusion, the study of Öztürk et al. (1) provides important data to the field of emergency medicine. However, a more in-depth analysis of critical variables such as treatment heterogeneity and a detailed classification of acidosis etiology would strengthen the clinical implications of the article and provide more concrete guidance to emergency physicians.

Best regards.

Corresponding Author: Sema Ayten MD, İstanbul Göztepe Prof. Dr. Süleyman Yalçın City Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye
E-mail: semayten@hotmail.com **ORCID ID:** orcid.org/0000-0001-8417-331X

Cite this article as: Ayten S. Management of patients with severe acidosis in the emergency department. Eurasian J Emerg Med. 2026;25: 261-2.

Received: 17.01.2026
Accepted: 24.03.2026
Published: 30.03.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Footnotes

Financial Disclosure: The author declared that this study received no financial support.

References

1. Öztürk. ZS. Outcome of cardiac arrest and non-cardiac arrest patients with severe acidosis in the emergency department: a retrospective cohort study. Eurasian Journal of Emergency Medicine. 2025;4:244-50.
2. Jaber S, Paugam C, Futier E, Lefrant JY, Lasocki S, Lescot T, et al. Sodium bicarbonate therapy for patients with severe metabolic acidaemia in the intensive care unit (BICAR-ICU): an open-label, randomised controlled, phase 3 trial. Lancet. 2018;392:31-40.
3. Jung B, Jabaudon M, De Jong A, Bitker L, Audard J, Klouche K, et al. Sodium Bicarbonate for Severe Metabolic Acidemia and Acute Kidney Injury: The BICARICU-2 Randomized Clinical Trial. JAMA. 2025;334:2000-10.
4. Bendiab E, Garnier F, Soler M, Fosset M, Jaber S, Molinari N, et al. Long-term outcome of severe metabolic acidemia in ICU Patients, a BICAR-ICU Trial Post Hoc Analysis. Crit Care Med. 2023;51:e1-e12.
5. Baddam S, Tubben RE. Lactic acidosis. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK470202/

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.66750
Eurasian J Emerg Med. 2026;25: 263-71

Knowledge, Attitudes, Self-efficacy, and Intention to Use Automated External Defibrillators Among Law Enforcement Officers: A Cross-sectional Survey

Medine Akkan Öz[1], Hüseyin Mutlu[2], Uğur Şakar[3], Yunus Yatmaz[4], Murat Tuğra Kösa[5], Müge Yenigün[6], Murat Genç[7], Ayşenur Gür[6], Esma Kır[1], Lukasz Szarpak[8]

[1]University of Health Sciences Türkiye, Ankara Gülhane Training and Research Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[2]Aksaray Training and Research Hospital, Clinic of Emergency Medicine, Aksaray, Türkiye
[3]Dışkapı Yıldırım Beyazıt Training and Research Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[4]Yenimahalle Training and Research Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[5]University of Health Sciences Türkiye, Ankara Etlik City Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[6]Etimesgut Şehit Sait Ertürk State Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[7]Ankara Training And Research Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[8]The John Paul Ii Catholic University of Lublin, Institute of Medical Science, Lublin Poland

Abstract

Aim: This study aimed to evaluate the knowledge, attitudes, perceived self-efficacy, intention to use automated external defibrillators (AEDs), and perceived barriers among law enforcement personnel (LEP) working within the Ankara Provincial Police Department. Additionally, the relationship between self-efficacy and intention to use AEDs was examined.

Materials and Methods: This single-center, cross-sectional study was conducted between January and February 2026 with 525 LEP. Data were collected via an anonymous, structured online questionnaire including sociodemographic items, a 10-item AED knowledge test, and Likert-type scales assessing attitudes, perceived self-efficacy, intention to use AEDs, and perceived barriers. Descriptive statistics, correlation, and regression analyses were performed.

Results: Participants demonstrated high AED knowledge and generally positive attitudes toward AED use. Perceived self-efficacy and intention-to-use scores indicated adequate readiness for intervention. A strong positive correlation was identified between self-efficacy and intention to use AEDs ($r=0.776$, $p<0.001$), with self-efficacy explaining a substantial proportion of the variance in intention. Personnel who had received AED-specific training showed significantly higher scores across all measured domains compared with untrained personnel. The most frequently reported barriers were lack of knowledge regarding AED location and uncertainty about device operation.

Conclusions: Although LEP showed favorable knowledge and attitudes toward AED use, AED-specific training and device accessibility remain essential for effective intervention. Perceived self-efficacy was the strongest predictor of intention to use an AED. Integrating targeted AED training, improving device visibility, and clarifying legal protections may enhance early defibrillation and improve out-of-hospital cardiac arrest outcomes.

Keywords: Automated external defibrillator, law enforcement personnel, self-efficacy, out-of-hospital cardiac arrest, first responder

Corresponding Author: Medine Akkan Öz MD, University of Health Sciences Türkiye, Ankara Gülhane Training and Research Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
E-mail: drakkanoz@gmail.com **ORCID ID:** orcid.org/0000-0002-6320-9667

Cite this article as: Akkan Öz M, Mutlu H, Şakar U, Yatmaz Y, Kösa MT, Yenigün M, et al. Knowledge, attitudes, self-efficacy, and intention to use automated external defibrillators among law enforcement officers: a cross-sectional survey. Eurasian J Emerg Med. 2026;25: 263-71.

Received: 26.02.2026
Accepted: 24.03.2026
Published: 30.03.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Introduction

Out-of-hospital cardiac arrest (OHCA) remains one of the leading causes of death worldwide, and the decisive impact of early recognition and timely intervention on survival is well established (1-3). The use of automated external defibrillators (AEDs), which constitute a critical link in the chain of survival, can substantially increase the likelihood of survival in shockable rhythms such as ventricular fibrillation and pulseless ventricular tachycardia, in which the probability of survival declines with each passing minute (2-5). In this context, considering that ambulance response times typically range from six to twelve minutes, the vital importance of early defibrillation performed by first responders at the scene becomes clearly evident (3,4). In the United States, the rate of cardiopulmonary resuscitation (CPR) initiation by police and fire personnel has been reported to be 31.8%, and survival was shown to increase by 1.4-fold in cases where an AED was applied (1). Similarly, a study conducted in Vienna evaluating a police-based AED first-responder system confirmed that police officers, with appropriate training, were able to provide high-quality basic life support and deliver effective intervention during the critical pre-ambulance interval (2).

Law enforcement personnel (LEP) occupy a strategic position as potential first responders in cases of OHCA due to their high visibility within the community, wide geographic distribution, and capacity for rapid response to emergencies (5-7). In many countries, AED devices have been installed in police vehicles, and LEP have been trained and authorized to provide first response care (2,7). However, in the United States, the level of knowledge regarding AED use was measured at only 31.9 out of 100, and merely 11.3% of LEP were found to have received AED training (3). In Europe, studies have reported low levels of first aid knowledge among LEP, noting that 61.7% of personnel encountered situations requiring intervention but did not know how to act. Additionally, it has been shown that LEP have insufficient knowledge of basic life support, with only about half of the personnel feeling competent in this regard (4,5). These findings underscore the need for targeted educational interventions directed at police populations.

Training programs specifically focused on AED use have been shown in various studies to positively influence LEP's knowledge, confidence, and intervention performance (6,7). In Amsterdam, a targeted three-hour training program provided to LEP increased their confidence in using an AED from 12% to 99% (6). Furthermore, in the United States, it has been reported that 96% of LEP with ten years of service agreed that AED use is beneficial for victims of cardiac arrest, and that these positive attitudes were sustained over the long term (7). These findings suggest that even short, targeted training interventions can produce lasting effects on self-efficacy and attitudes. However, a substantial gap persists between knowledge and actual intervention behavior, even among trained individuals (8). Behavioral science theories offer an important framework for explaining this knowledge–behavior gap (9). Rather than focusing solely on knowledge transfer, a CPR performance model centered on intention proposes that self-efficacy and behavioral intention are the primary determinants of intervention behavior (9). According to the Theory of Planned Behavior, attitudes, subjective norms, and perceived behavioral control (self-efficacy) jointly shape behavioral intention, which in turn influences actual behavior. Large-sample studies have validated this theoretical framework and demonstrated that perceived behavioral control is a significant predictor of the intention to use an AED (10,11). Moreover, several studies have shown that training can significantly increase confidence in AED use, thereby confirming that self-efficacy is a modifiable construct (11,12).

To date, no study in Türkiye has comprehensively evaluated LEP's knowledge, attitudes, self-efficacy, and intention to use AEDs using quantitative scales. This study aims to assess the level of knowledge, attitudes, perceived self-efficacy, intention to use AEDs, and perceived barriers among LEP working within the Ankara Provincial Police Department; to examine the relationship between self-efficacy and intention to use AEDs; and to identify the sociodemographic and occupational factors associated with self-efficacy and intention to use AEDs. The study hypothesis is that higher perceived self-efficacy regarding AED use will be associated with a higher intention to use an AED among LEP.

Materials and Methods

Study Design and Participants

This study was designed as a single-center, cross-sectional, descriptive, and analytical survey to evaluate the knowledge level, attitudes, perceived self-efficacy, intention to use AEDs, and perceived barriers among LEP working within the Ankara Provincial Police Department. The study was approved by the Non-Interventional Clinical Research Ethics Committee of the Ankara Provincial Health Directorate of the Ministry of Health (approval no: 2025-12-7, date: 30.12.2025). The research was conducted in accordance with the principles of the Declaration of Helsinki. The study was carried out between January and February 2026 at the First Aid Training Unit affiliated with the Ankara Provincial Health Directorate. The first page of the online questionnaire included a participant information text and an electronic informed consent statement. Participation was entirely voluntary, and informed consent was obtained from all

participants before they proceeded to the survey. LEP aged 18 years and older who were actively working within the Ankara Provincial Police Department and who voluntarily agreed to participate in the study were eligible for inclusion. Participants who refused to complete the questionnaire, wished to withdraw during data collection, or left more than 50% of the questionnaire unanswered were excluded from the study.

Data Collection Instrument

A self-report, structured questionnaire developed by the researchers based on relevant literature was used as the data collection tool. The questionnaire was distributed online through Google Forms. No personally identifiable information was collected, and responses were analyzed anonymously. Participants were informed about the study's purpose and told that their responses would be used solely for research purposes and that no individual evaluation would take place. They were also told they could leave any question blank or withdraw at any time. The dataset was stored securely and only accessible to the research team.

The questionnaire consisted of two main sections:

Section 1: Sociodemographic and occupational characteristics. This section included a total of 12 items assessing participants' age, sex, educational level, duty unit, years of service, work schedule, prior first aid training and the timing of the most recent training, previous AED-specific training, experience of encountering or witnessing a cardiac arrest event, performance of CPR during duty, prior exposure to an AED device, and level of knowledge regarding the location of AEDs within their duty area. In this study, "first aid training" and "AED-specific training" referred to training that participant had previously received; no educational intervention was delivered as part of the study.

Section 2: AED knowledge test and Likert-type scales. This section comprised five subdimensions:

Knowledge test: This component consisted of 10 multiple-choice questions addressing the indications for AED use, steps of application, safety precautions, and integration with basic life support. Each question provided five options, of which only one was accepted as the correct answer. Correct responses were coded as 1 and incorrect responses as 0, and the total knowledge score was calculated on a scale ranging from 0 to 10. Knowledge levels were categorized as high (8-10 points), moderate (5-7 points), and low (0-4 points).

Attitude scale: This scale consisted of 8 items assessing attitudes toward AED use. The items were rated on a 5-point Likert scale (1=strongly disagree, 5=strongly agree). Item 2 ("AED use is solely the responsibility of healthcare personnel"), item 4 ("using an AED is unnecessary; it is better to wait until the ambulance arrives"), and item 7 ("using an AED interferes with my duties at the scene") were negatively worded and were reverse-coded during analysis. The mean item score was calculated, with higher scores interpreted as reflecting a more positive attitude.

Self-efficacy scale: This scale consisted of 6 items measuring perceived self-efficacy related to AED use. The items assessed competencies including recognizing the need for an AED in an unconscious patient, turning on the device and following voice prompts, correctly placing the pads, ensuring scene safety, using the device under stress, and coordinating role allocation within the team. Responses were rated on a 5-point Likert scale, with higher scores indicating greater perceived self-efficacy.

Intention to use scale: This scale consisted of 4 items evaluating the intention and willingness to use an AED. The items measured willingness to use an AED when appropriate, efforts to learn the device location, intention to encourage colleagues to use the device, and determination to participate in the intervention without hesitation. Responses were rated on a 5-point Likert scale.

Perceived barriers scale: This scale consisted of 10 items assessing perceived barriers that may hinder AED use. The items covered dimensions including fear of incorrect application and causing harm, concerns about legal liability, fear of infection transmission, difficulty providing medical intervention under potential safety risks at the scene, lack of knowledge about AED location, tendency to wait for healthcare teams, fear of making mistakes under crowd pressure, uncertainty about the operating principles of the AED, the influence of whether the victim is familiar or unfamiliar on the decision to intervene, and perceiving AED use as outside the scope of duty. Responses were rated on a 5-point Likert scale, with higher scores indicating greater perceived barriers.

The total number of items in the questionnaire, including the sociodemographic section, was 50. The internal consistency reliability of the Likert-type scales was evaluated using Cronbach's alpha coefficient. Cronbach's alpha was calculated as 0.810 for the attitude scale, 0.907 for the self-efficacy scale, 0.897 for the intention to use scale, and 0.785 for the perceived barriers scale. All values were above the acceptable threshold for reliability.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics version 26.0 (IBM Corp., Armonk, NY, USA). A p-value of <0.05 was considered statistically significant for all analyses. For descriptive statistics, continuous variables were presented as mean $\pm$ standard deviation, median, and minimum-maximum values,

whereas categorical variables were expressed as frequency (n) and percentage (%). The normality of continuous variables was assessed using the Shapiro-Wilk and Kolmogorov-Smirnov tests. When the assumption of normal distribution was not met, both parametric and non-parametric tests were applied and the consistency of the results was verified. The internal consistency reliability of the scales was evaluated using Cronbach's alpha coefficient, with values of ≥0.70 considered indicative of acceptable reliability. For comparisons between two independent groups (AED training status, prior first aid training, sex, and witnessing a cardiac arrest), the independent samples t-test and the Mann-Whitney U test were used. For comparisons involving more than two independent groups (years of service, educational level, and duty unit), ANOVA and the Kruskal-Wallis H test were employed. Pearson and Spearman correlation analyses were conducted to examine relationships among the main variables. Correlation coefficients were interpreted as follows: 0.10-0.29, weak; 0.30-0.49, moderate; 0.50-0.69, strong; and ≥0.70, very strong correlation. Binary logistic regression was performed to identify factors associated with high intention to use an AED. For this analysis, the intention-to-use score was dichotomized as high intention (mean score ≥4.0) versus lower intention (mean score <4.0). Age, sex, prior first aid training, prior AED-specific training, witnessing a cardiac arrest, prior exposure to an AED device, knowledge of AED location, knowledge score, attitude score, self-efficacy score, and perceived barriers score were entered into the model. Odds ratios (ORs) with 95% confidence intervals (CIs) were reported.

Results

A total of 525 LEP participated in the study. The mean age of the participants was 36.9±8.5 years, and 78.9% (n=414) were male. Regarding educational level, 66.5% (n=352) held a bachelor's degree and 18.9% (n=100) had an associate degree. In terms of years of service, the largest group consisted of personnel with 6-10 years of experience (23.1%). It was found that 68.8% (n=361) of the personnel had previously received first aid training; however, only 15.5% (n=82) had received AED-specific training. A total of 18.5% (n=98) reported having previously encountered a cardiac arrest event, while only 3.2% (n=17) had performed CPR during duty. The proportion of participants who reported having previously seen an AED device was 29.9% (n=158), and 19.5% (n=103) stated that they knew the location of an AED in their duty area; 21.7% (n=115) were unsure of the location, and 58.0% (n=307) reported that they did not know. Detailed sociodemographic characteristics of the participants are presented in Table 1.

Table 1. Sociodemographic characteristics of the participants (n=525)

Variable	Category	n	%
Age, years	Mean ± SD	36.9±8.5	
	Median (min-max)	37 (20-68)	
Age group, years	18-25	49	9.3
	26-30	66	12.6
	31-35	121	23
	36-40	126	24
	41-45	76	14.5
	46-50	54	10.3
	≥51	33	6.3
Sex	Male	414	78.9
	Female	106	20.2
	Prefer not to specify	5	0.9
Educational level	High school	11	2.1
	Associate degree	100	19
	Bachelor's degree	352	67.1
	Postgraduate degree	62	11.8
Years of service	0-5	104	19.8
	6-10	122	23.2
	11-15	76	14.5
	16-20	102	19.4
	21-25	45	8.6
	>25	76	14.5
Duty unit	Special operations unit	141	26.9
	Office/administrative	115	21.9
	Riot police	58	11.0
	Public order/police station	55	10.4
	Narcotics/protection/intelligence	46	8.8
	Traffic/patrol	34	6.5
	Other	76	14.5
Work schedule	Day shift	248	47.2
	Variable shift	155	29.5
	Shift work	122	23.2
First aid training	Received	361	68.8
	Not received	164	31.2
AED training	Yes	82	15.6
	No	443	84.4
Witnessing a cardiac arrest	Yes	98	18.6
	No	427	81.3
CPR performed during duty	Yes	17	3.2
	No	508	96.8
Prior exposure to an AED device	Yes	158	30.1
	No	367	69.9
Knowledge of AED location	Yes	103	19.6
	No	307	58.5
	Not sure	115	21.9

SD: Standard deviation, AED: Automated external defibrillator, CPR: Cardiopulmonary resuscitation

The mean number of correct responses in the AED knowledge test among LEP was 8.49±1.64 (84.9% out of 100). When knowledge levels were categorized, 81.3% of participants (n=427) were found to have a high level of knowledge (8-10 correct answers). For the attitude scale (8 items, Cronbach's α=0.810), the mean item score was 4.08±0.58. The highest attitude score was observed for the item "when used correctly, AED use is a life-saving intervention in the field" (4.25±0.89), whereas the lowest score was recorded for the reverse-coded item "AED use is solely the responsibility of healthcare personnel" (3.89±0.92). These findings indicate that LEP generally have a positive attitude toward AED use. On the self-efficacy scale (6 items, Cronbach's α=0.907), the mean item score was 3.77±0.77. The highest self-efficacy score was observed for the item "I can ensure scene safety and confirm that no one is in contact before delivering a shock" (4.14±0.89), while the lowest score was obtained for the item "I can recognize the need for an AED in a person who is unconscious and not breathing normally" (3.42±0.98). For the perceived barriers scale (10 items, Cronbach's α=0.785), the mean item score was calculated as 3.04±0.55. All scale scores and detailed subdimension results are summarized in Table 2.

Table 2. Descriptive statistics of the scales and subdimension analysis (n=525)

Subdimension/item	Mean	SD	Min-Max	α
Knowledge test (0-10 points)	8.49	1.64	0-10	
Percentage of correct responses	84.9%			
High knowledge (8-10 points)	n=427	81.3%		
Moderate knowledge (5-7 points)	n=84	16%		
Low knowledge (0-4 points)	n=14	2.7%		
Attitude scale (mean item score)	4.08	0.58	1-5	0.810
Having basic knowledge is necessary	4.05	0.92	1-5	
It is the responsibility of healthcare personnel (R)	3.89	0.92	1-5	
Voice prompts facilitate use	4.14	0.86	1-5	
Unnecessary; wait for the ambulance (R)	4.11	0.86	1-5	
Visibility in public areas	4.17	0.88	1-5	
It is a life-saving intervention	4.25	0.89	1-5	
It interferes with my duties (R)	3.92	0.90	1-5	
Proper use can be taught	4.08	0.88	1-5	
Self-efficacy scale (mean item score)	3.77	0.77	1-5	0.907
I can recognize the need for an AED	3.42	0.98	1-5	
I can follow voice prompts	3.78	0.93	1-5	
I can correctly place the pads	3.85	0.95	1-5	
I can ensure scene safety	4.14	0.89	1-5	
I can use the device under stress	3.68	0.92	1-5	
I can coordinate role allocation within the team	3.74	0.95	1-5	
Intention to use scale (mean item score)	3.74	0.83	1-5	0.897
I am willing to use an AED	3.71	0.90	1-5	
I try to learn the device location	3.66	0.97	1-5	
I encourage my colleagues	3.77	0.94	1-5	
I intervene without hesitation	3.81	0.98	1-5	
Perceived barriers (mean item score)	3.04	0.55	1-5	0.785
Fear of making a mistake and causing harm	2.90	0.95	1-5	
Concerns about legal liability	3.16	0.95	1-5	
Fear of infection transmission	2.85	0.94	1-5	
Providing medical intervention under safety risk	3.33	0.96	1-5	
Not knowing the AED location	3.81	0.92	1-5	
Belief that waiting for the healthcare team is appropriate	2.63	0.94	1-5	
Fear of making mistakes due to crowd pressure	2.79	0.93	1-5	
Uncertainty about the operating principles of the AED	3.76	0.88	1-5	
Influence of familiarity with the victim (known vs unknown)	2.44	0.98	1-5	
Feeling that AED use is outside the scope of duty	2.73	0.97	1-5	

T: Reverse-coded item, α: Cronbach's alpha reliability coefficient, SD: Standard deviation, AED: Automated external defibrillator

A statistically significant, strong, and positive correlation was found between self-efficacy and intention to use AEDs (Pearson r=0.776; Spearman rs=0.779; p<0.001 for both). Self-efficacy alone explained 60.3% of the variance in intention to use (R^2=0.603). In the simple linear regression analysis, higher self-efficacy scores were associated with higher intention-to-use scores; specifically, each one-point higher self-efficacy score corresponded to a 0.832-point higher intention-to-use score (β=0.832; standard error=0.030; t=28.173; p<0.001). The Pearson correlation matrix among all main variables is presented in Table 3.

In the comparison based on AED training status, personnel who had previously received AED training (n=82) demonstrated significantly higher knowledge scores (9.35±1.15 vs 8.40±1.48; t=5.54; p<0.001), attitude scores (4.30±0.54 vs 4.03±0.58; t=3.94; p<0.001), self-efficacy scores (4.07±0.70 vs 3.71±0.78; t=3.85; p<0.001), and intention to use scores (4.06±0.72 vs 3.68±0.84; t=3.87; p<0.001) compared with those who had not received AED training. When evaluated according to prior first aid training status, no statistically significant differences were observed between trained and untrained groups in terms of knowledge scores (8.56±1.73 vs 8.31±1.43; t=1.64; p=0.102), self-efficacy scores (3.80±0.76 vs 3.71±0.81; t=1.17; p=0.240), or intention to use scores (3.76±0.81 vs 3.69±0.87; t=0.87; p=0.384). These findings suggest that prior general first aid training may not be associated with better AED-specific knowledge, self-efficacy, or intention-to-use scores in this sample.

Binary logistic regression was performed to identify factors associated with high intention to use an AED. High intention was defined as a mean intention-to-use score of ≥4.0. In univariable analyses, prior AED-specific training, knowledge score, attitude score, self-efficacy score, and perceived barriers score were significantly associated with high intention. In the multivariable model, higher self-efficacy scores remained strongly associated with high intention to use an AED (adjusted OR=11.00, 95% CI=6.35-19.08, p<0.001). More favorable attitudes were also independently associated with high intention (adjusted OR=3.32, 95% CI=1.68-6.55, p=0.001), whereas higher perceived barriers were independently associated with lower odds of high intention (adjusted OR=0.44, 95% CI=0.24-0.80, p=0.007). Age, sex, prior first aid training, prior AED-specific training, witnessing a cardiac arrest, prior exposure to an AED device, knowledge of AED location, and knowledge score were not independently associated with high intention after adjustment (Table 4).

Table 3. Pearson correlation matrix among the main variables

	Knowledge	Attitude	Self-efficacy	Intention	Barriers
Knowledge	1	0.332***	0.230***	0.260***	-0.105*
Attitude		1	0.690***	0.722***	-0.281***
Self-efficacy			1	0.776***	-0.291***
Intention				1	-0.345***
Barriers					1

*p<0.05, **p<0.01, ***p<0.001

Table 4. Univariable and multivariable binary logistic regression analyses of factors associated with high intention to use an automated external defibrillator

Variables	Univariate logistic regression		Multivariate logistic regression	
	OR (95% CI)	p	OR (95% CI)	p
Age	0.98 (0.96-1.00)	0.135	0.99 (0.96-1.02)	0.500
Female sex	0.85 (0.55-1.31)	0.468	0.96 (0.53-1.75)	0.906
Prior first aid training	1.21 (0.84-1.76)	0.304	1.27 (0.76-2.12)	0.366
Prior AED-specific training	1.87 (1.16-3.02)	0.010	0.90 (0.46-1.76)	0.761
Witnessing a cardiac arrest	0.92 (0.59-1.42)	0.697	0.86 (0.47-1.57)	0.628
Prior exposure to an AED device	1.35 (0.93-1.96)	0.115	0.93 (0.55-1.56)	0.772
Knowledge of AED location	1.40 (0.91-2.16)	0.128	0.91 (0.51-1.64)	0.762
Knowledge score	1.24 (1.09-1.42)	0.001	1.01 (0.83-1.23)	0.899
Attitude score	16.01 (9.19-27.88)	<0.001	3.32 (1.68-6.55)	0.001
Self-efficacy score	19.90 (11.86-33.41)	<0.001	11.00 (6.35-19.08)	<0.001
Perceived barriers score	0.32 (0.22-0.48)	<0.001	0.44 (0.24-0.80)	0.007

High intention was defined as a mean intention-to-use score of ≥4.0. In the multivariable model, female sex was compared with male sex, and knowledge of AED location was coded as yes versus no/not sure. The multivariable model included 520 complete cases
OR: Odds ratio, CI: Confidence interval, AED: Automated external defibrillator

Discussion

Worldwide, CPR and AED training is provided to the general public and various professional groups to enable early recognition of OHCA, ensure timely intervention, and ultimately improve survival rates; increasing the availability of such training is strongly recommended (1,7). However, debates persist regarding the extent to which the knowledge acquired during training is effectively translated into real-life practice (5,9). In this study, we evaluated knowledge, attitudes, perceived self-efficacy, and intention to use AEDs among LEP in Türkiye. To the best of our knowledge, this is the first study on this topic conducted in Türkiye. Overall, the findings suggest that participants had generally favorable levels of knowledge, attitudes, and willingness regarding AED use, and that prior AED-specific training was associated with more favorable outcomes across multiple domains. In addition, prior AED-specific training was associated with more favorable scores across multiple domains, and self-efficacy showed a strong positive association with intention to use an AED. In light of these findings, AED-specific training that includes simulation-based components and periodic refresher sessions may help strengthen preparedness and willingness to intervene among LEP.

In a cross-sectional study conducted by Mohd Hashim et al. (11) in Malaysia with 217 community volunteers, self-efficacy was reported to be among the most influential factors determining willingness to perform CPR and use an AED. Similarly, Liaw et al. (12) reported that improved confidence regarding AED use was accompanied by greater willingness to use the device, while Gao et al. (13) demonstrated that higher self-efficacy was associated with improved CPR-related knowledge, attitudes, and practices with these studies, our findings also showed a strong positive association between self-efficacy and intention to use an AED.

In our study, the knowledge level regarding AED use appeared comparatively high. This proportion is markedly higher than the values reported for police populations in the international literature. In a reference study conducted by Groh et al. (3) in the United States with 929 law enforcement officers, the mean knowledge score was only 31.9 out of 100, and only 11.3% of the personnel had received AED training. In Greece, Tzenetidis et al. (4) found that the correct response rate to first aid and CPR questions among 520 police officers was 34.8%, while the proportion who had received AED-specific training was only 6.5%. In Spain, a study conducted with local police officers by Angulo-Menéndez et al. (5) reported that only 11% correctly identified the appropriate chest compression depth and 24.4% correctly identified the appropriate compression rate. Although the comparatively high level of knowledge observed in our sample may partly reflect the growing emphasis on first aid training in law enforcement training programs in recent years, this finding should nevertheless be interpreted with caution. Several methodological factors may have contributed to the high scores observed. First, because the survey was administered online, it was not possible to fully control whether participants consulted external resources while answering the knowledge items. Second, some of the questions may have addressed relatively basic or easily recognizable aspects of AED use, thereby increasing the likelihood of correct responses. Third, given the voluntary nature of participation, selection bias cannot be excluded; individuals who were more motivated, more confident, or more interested in first aid-related issues may have been more likely to participate. In addition, the multiple-choice format itself may have facilitated correct responses independent of actual procedural competence. Taken together, these considerations suggest that the knowledge scores reported here may reflect performance under survey conditions rather than actual field readiness.

Our findings suggest that LEP generally held favorable attitudes toward AED use. In a study by Papson and Mosesso (7) evaluating ten-year police AED programs in the United States, 96% of personnel acknowledged that police use of AEDs benefits victims of cardiac arrest, and 89% considered this practice appropriate. In a national survey conducted in the United Kingdom with 2,084 participants, Hawkes et al. (14) also found that prior training experience was most strongly associated with willingness to use an AED. Consistent with previous studies, personnel with prior AED training in our sample also exhibited more favorable attitudes toward AED use. This consistency is in keeping with an association between AED-related training exposure and more favorable attitudes toward AED use.

One notable finding of our study was that prior AED-specific training was associated with more favorable outcomes across several domains, whereas prior general first aid training alone did not show the same pattern. This finding may indicate that prior general first aid training alone is not necessarily associated with stronger AED-specific competencies. In a landmark study conducted with the Amsterdam police, Kooij et al. (6) demonstrated that following a three-hour AED-specific training program, personnel's confidence in using the device increased from 12% to 99%, while their motivation rose from 73% to 94%. This dramatic improvement indicates that even short, targeted training interventions can have a strong impact on self-efficacy.

In the perceived barriers domain, the most prominent concerns were related to lack of knowledge about AED location and uncertainty regarding device operation. These findings suggest that reluctance to intervene may be more closely associated with infrastructural and informational barriers than with individual

fear or attitudinal resistance. In a nationwide telephone survey conducted in Taiwan, Huang et al. (15) reported that only 15.2% of the public were able to recognize public AED devices and only 5.3% knew how to use them. Moreover, although willingness to use an AED was high (86.6%), only 37.3% of participants were able to identify the location of a nearby AED (16). In our sample as well, awareness of AED location within the duty area was limited, supporting the view that accessibility and visibility remain important practical barriers. Similarly, in a cross-sectional study conducted in Hong Kong, Chow demonstrated that unguided AED placement and concerns about legal liability were among the primary barriers (17). Concerns about legal liability were also among the barriers identified in our sample.

When the overall pattern of relationships was considered, attitude and self-efficacy were positively associated with the intention to use an AED, whereas perceived barriers were inversely associated with these constructs. This pattern is consistent with the network of relationships proposed by the Theory of Planned Behavior (9) and parallels the knowledge–attitude correlation (r=0.782) reported in Chow's study conducted in Hong Kong (17). Given the observed associations among these variables, intervention programs may benefit from addressing perceived barriers in addition to improving knowledge and skills.

In the logistic regression analysis, self-efficacy emerged as the variable most strongly associated with high intention to use an AED after adjustment for demographic, training-related, and other scale variables. Attitude also remained independently associated with high intention, whereas perceived barriers showed an inverse association. Taken together, these findings suggest that willingness to use an AED may be more closely related to confidence, attitudinal orientation, and perceived obstacles than to demographic characteristics alone.

One of the strengths of our study is the relatively large sample size (n=525 LEP). However, several situations should be acknowledged. First, the study was designed as a cross-sectional investigation; therefore, causal relationships between variables cannot be established, and only associative inferences can be made. Second, the study was conducted in a single province (Ankara), and thus caution should be exercised when generalizing the findings to all LEP. Finally, due to the online administration of the questionnaire, it was not possible to fully control whether participants consulted external sources while responding. Overall, our findings suggest that Turkish LEP have a generally favorable orientation toward AED use, although important gaps remain regarding AED-specific training and practical readiness.

Study Limitations

Nevertheless, limited AED-specific training and restricted awareness of device accessibility appear to be associated with an important gap between favorable attitudes and readiness to intervene.

Conclusion

In this cross-sectional study, LEP generally demonstrated positive knowledge, attitudes, and willingness regarding AED use. Among the measured aspects, perceived self-efficacy showed the strongest link to the intention to use an AED. Prior AED-specific training and limited awareness of device locations were identified as clinically relevant findings that could inform future training and implementation strategies. Incorporating AED-specific, simulation-based training into routine first aid programs and enhancing the visibility and accessibility of AEDs in duty areas may help improve preparedness for early intervention. Future multicenter, interventional studies are needed to further explore these relationships and assess the impact of targeted educational strategies.

Ethics

Ethics Committee Approval: Ankara Provincial Police Department. The study was approved by the Non-Interventional Clinical Research Ethics Committee of the Ankara Provincial Health Directorate of the Ministry of Health (approval no: 2025-12-7, date: 30.12.2025).

Informed Consent: This single-center, cross-sectional study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: M.A.Ö., H.M., Y.Y., M.T.K., M.Y., M.G., E.K., Concept: M.A.Ö., H.M., U.Ş., Y.Y., M.Y., M.G., E.K., Design: M.A.Ö., H.M., U.Ş., M.T.K., M.G., A.G., E.K., L.S., Data Collection or Processing: M.A.Ö., H.M., U.Ş., Y.Y., M.T.K., A.G., L.S., Analysis or Interpretation: M.A.Ö., H.M., Y.Y., M.T.K., M.Y., M.G., L.S., Literature Search: M.A.Ö., H.M., Y.Y., M.T.K., M.Y., M.G., L.S., Writing: M.A.Ö., H.M., U.Ş., Y.Y., M.T.K., M.G., L.S.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Salhi RA, Hammond S, Lehrich JL, O'leary M, Kamdar N, Brent C, et al. The association of fire or police first responder initiated interventions with out of hospital cardiac arrest survival. Resuscitation. 2022;174:9-15. Epub 2022 Mar 4.
2. Krammel M, Lobmeyr E, Sulzgruber P, Winnisch M, Weidenauer D, Poppe M, et al. The impact of a high-quality basic life support police-based first responder system on outcome after out-of-hospital cardiac arrest. PLoS One. 2020;15:e0233966.
3. Groh WJ, Lowe MR, Overgaard AD, Neal JM, Fishburn WC, Zipes DP. Attitudes of law enforcement officers regarding automated external defibrillators. Acad Emerg Med. 2002;9:751-3.
4. Tzenetidis N, Kourlaba G, Triantafyllou C, Tzenetidis V, Koulouvaris P, Myrianthefs P. Practices and educational needs of Hellenic Police Officers in providing first aid/cardiopulmonary resuscitation: A cross-sectional survey. Work. 2023;75:975-86.
5. Angulo-Menéndez P, Lana Pérez A, Morís de la Tassa J. Conocimientos y disposición para realizar soporte vital básico por agentes de la policía local. An Sist Sanit Navar. 2017;40:177-85.
6. Kooij FO, van Alem AP, Koster RW, de Vos R. Training of police officers as first responders with an automated external defibrillator. Resuscitation. 2004;63:33-41.
7. Papson K, Mosesso VN Jr. Ten years of police defibrillation: program characteristics and personnel attitudes. Prehosp Emerg Care. 2005;9:186-90.
8. Özbilgin Ş, Akan M, Hancı V, Aygün C, Kuvaki B. Evaluation of public awareness, knowledge and attitudes about cardiopulmonary resuscitation: Report of İzmir. Turk J Anaesthesiol Reanim. 2015;43:396-405.
9. Regard S, Rosa D, Suppan M, Giangaspero C, Larribau R, Niquille M,et al. Evolution of Bystander Intention to Perform Resuscitation Since Last Training: Web-Based Survey. JMIR Form Res. 2020;4:e24798.
10. Tam VCW, Kwok AWL. Evaluation of the willingness to perform cardiopulmonary resuscitation (CPR) with automated external defibrillator (AED) among Hong Kong Chinese using the theory of planned behaviour framework: an online cross-sectional survey. BMJ Public Health. 2024;2:e000406.
11. Mohd Hashim NI, Daud A, Mohammed Nawi A. Factors influencing willingness to perform cardiopulmonary resuscitation (CPR) and use an automated external defibrillator (AED) among non-healthcare community participants in a CPR fun run. BMC Public Health. 2025;25:2944.
12. Liaw SY, Chew KS, Zulkarnain A, Wong SSL, Singmamae N, Kaushal DN, et al. Improving perception and confidence towards bystander cardiopulmonary resuscitation and public access automated external defibrillator program: how does training program help? Int J Emerg Med. 2020;13:13.
13. Gao H, Liu X, Jiang Z, Huang S, Pan X, Long J, et al. Knowledge, attitudes, practices, and self-efficacy of the Chinese public regarding cardiopulmonary resuscitation: an online cross-sectional survey. Front Public Health. 2024;12:1341851.
14. Hawkes CA, Brown TP, Booth S, Fothergill RT, Siriwardena N, Zakaria S, et al. Attitudes to Cardiopulmonary Resuscitation and Defibrillator Use: A Survey of UK Adults in 2017. J Am Heart Assoc. 2019;8:e008267.
15. Huang EPC, Chiang WC, Lu TC, Wang CH, Sun JT, Hsieh MJ, et al. Barriers to bystanders defibrillation: A national survey on public awareness and willingness of bystanders defibrillation. J Formos Med Assoc. 2021;120:974-82.
16. Huang CT, Chen CH, Huang CH, Fan CY, Chen JW, Ma MH, et al. Public awareness of automated external defibrillator locations. JAMA Netw Open. 2024;7:e2438319.
17. Chow SKY. Bystanders' Views on the use of automated external defibrillators for out-of-hospital cardiac arrest: implications for health promotions. International Journal of Environmental Research and Public Health. 2021;18:1241.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Letter to the Editor
DOI: 10.4274/eajem.galenos.2026.95825
Eurasian J Emerg Med. 2026;25: 272-4

Multidisciplinary AI-assisted Diagnosis and Management of Acute Abdominal Pain Syndrome: A Pilot Study

Mete Üçdal[1], Evren Ekingen[2]

[1]Etimesgut Şehit Sait Ertürk State Hospital, Clinic of Internal Medicine, Ankara, Türkiye
[2]Etimesgut Şehit Sait Ertürk State Hospital, Clinic of Emergency Medicine, Ankara, Türkiye

Dear Editor,

Acute abdominal pain syndrome represents one of the most common and diagnostically challenging presentations in emergency medicine, accounting for approximately 5-10% of all emergency department visits (1). The differential diagnosis spans diverse surgical emergencies including acute appendicitis, acute cholecystitis, intestinal obstruction, and volvulus, each requiring distinct management strategies ranging from conservative therapy to emergent surgical intervention (2). Diagnostic accuracy is paramount, as delayed recognition of surgical emergencies significantly increases morbidity and mortality. Recent advances in multimodal artificial intelligence (AI) systems have demonstrated promise in clinical decision support; however, their application across the heterogeneous spectrum of acute abdominal emergencies remains limited (3). We present a pilot evaluation comparing a multidisciplinary AI system with a single-model AI for diagnosing and managing acute abdominal pain syndrome.

We retrospectively analyzed 36 consecutive patients who presented to Etimesgut Şehit Sait Ertürk State Hospital (Ankara, Türkiye) with acute abdominal pain between September 2024 and January 2025. A multidisciplinary AI system integrating GPT-4V for radiological interpretation (ultrasonography, computed tomography findings), Med-PaLM 2 for emergency medicine assessment, BioGPT for intensive care unit (ICU) triage, and LLaMA 4 for surgical decision-making was compared against ChatGPT-5.0 and decisions made by the clinical team. GPT-4V was prompted using standardized radiological descriptor templates to identify pathognomonic imaging features: appendiceal diameter ≥6 mm with periappendiceal fat stranding (appendicitis); gallbladder wall thickening ≥3 mm, pericholecystic fluid, and sonographic Murphy sign (cholecystitis); and whirl sign, coffee-bean sign, and transition point (volvulus). Med-PaLM 2 processed structured clinical parameters including vital signs, symptom onset, physical examination findings (rebound tenderness, guarding, McBurney's tenderness), and laboratory values (white blood cell, C-reactive protein, bilirubin, lipase, lactate) using PICO-formatted prompts. BioGPT assessed the necessity for ICU admission based on Sequential Organ Failure Assessment scores, the systemic inflammatory response syndrome criteria, hemodynamic parameters, and lactate levels according to the Surviving Sepsis Campaign 2023 guidelines. LLaMA 4 integrated American Society of Anesthesiologists physical status classification, comorbidity profile, and current guideline recommendations [World Society of Emergency Surgery (WSES), Tokyo Guidelines 2018, American Society for Gastrointestinal Endoscopy (ASGE)] to generate urgency-stratified surgical management decisions. All AI models, which did not have access to clinical team decisions, received standardized, pre-specified prompts that were developed a priori and applied uniformly across all 36 cases. Final diagnoses, established by consensus of the attending clinical team (emergency physician, surgeon, and radiologist), were confirmed by operative findings, pathological reports, or clinical course as appropriate, and served as the reference standard for all AI performance evaluations. AI model assessments were performed independently and were blinded to the clinical team's decisions, using standardized, pre-specified prompts applied uniformly across all cases. The cohort (mean age 52.4±17.6 years; 55.6% male) comprised acute appendicitis (n=12, 33.3%), acute cholecystitis (n=10, 27.8%), paralytic ileus

Corresponding Author: Mete Üçdal MD, Etimesgut Şehit Sait Ertürk State Hospital, Clinic of Internal Medicine, Ankara, Türkiye
E-mail: meteucdal@hacettepe.edu.tr **ORCID ID:** orcid.org/0000-0002-2321-2642
Cite this article as: Üçdal M, Ekingen E. Multidisciplinary AI-assisted diagnosis and management of acute abdominal pain syndrome: a pilot study. Eurasian J Emerg Med. 2026;25: 272-4.

Received: 09.01.2026
Accepted: 30.03.2026
Published: 01.04.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

(n=8, 22.2%), sigmoid volvulus (n=4, 11.1%), and cecal volvulus (n=2, 5.6%).

The multidisciplinary AI achieved 94.4% overall diagnostic accuracy (34/36), significantly outperforming ChatGPT-5.0 (80.6%, 29/36; p=0.012). For acute appendicitis, the multidisciplinary AI system correctly identified 11/12 cases (91.7%), including 3 cases of complicated appendicitis with perforation, compared with 9/12 (75.0%) identified by ChatGPT-5.0; integration of the Alvarado score further enhanced diagnostic precision. Detection of acute cholecystitis achieved 90.0% sensitivity (9/10) by accurately identifying gallbladder wall thickening and pericholecystic fluid on ultrasonography. The critical differentiation between surgical emergencies (appendicitis, cholecystitis, volvulus) and candidates for conservative management (ileus) showed 96.4% accuracy (27/28) for the multidisciplinary system, versus 82.1% (23/28) for ChatGPT-5.0 (p=0.024). Volvulus cases demonstrated a detection sensitivity of 100% with recognition of the coffee-bean and whirl signs (Table 1).

Analysis of guideline adherence revealed concordance rates of 94.4% (34/36) for multidisciplinary AI compared with 77.8% (28/36) for ChatGPT-5.0 (p=0.018). For appendicitis, the AI correctly recommended early laparoscopic appendectomy per WSES Jerusalem guidelines in all uncomplicated cases (4). Acute cholecystitis management was aligned with Tokyo Guidelines 2018 criteria, with appropriate early cholecystectomy recommended in 9/10 cases (90.0%) versus 7/10 (70.0%) for ChatGPT-5.0 (5). Sigmoid volvulus cases received guideline-concordant endoscopic detorsion recommendations (100% vs 50.0%) per ASGE criteria (6). ICU triage accuracy was 91.7% versus 77.8% (p=0.032), with BioGPT demonstrating superior sepsis recognition per Surviving Sepsis Campaign criteria (7). Clinical outcomes were: surgical intervention in 22 patients (61.1%), endoscopic intervention in 3 patients (8.3%), and conservative management in 11 patients (30.6%); there was one in-hospital death (perforated appendicitis, 2.8%) (Table 2).

Our findings support the emerging paradigm of task-distributed AI in acute care. Schouten et al. (8) reported in a comprehensive scoping review analyzing 432 studies that multimodal AI models provide an average 6.2-point area under the curve improvement over unimodal systems, consistent with our 13.8 percentage point accuracy difference. Notably, the 94.4% diagnostic accuracy achieved by the multidisciplinary AI system is contextually comparable to published benchmarks: Kaczmarczyk et al. (9) reported 95.8% accuracy for collective human decision-making in complex diagnostic tasks; however, given the retrospective single-center design of the current pilot study, direct equivalence with human specialist performance cannot be inferred and

Table 1. Diagnostic performance of multidisciplinary ai across acute abdominal pain subtypes

Condition	n (%)	Multidisc. AI	ChatGPT-5.0	Sens. (%)	Key finding
Acute appendicitis	12 (33.3)	11/12	9/12	91.7 vs 75.0	Alvarado score
Acute cholecystitis	10 (27.8)	9/10	7/10	90.0 vs 70.0	Wall thickening
Paralytic ileus	8 (22.2)	8/8	7/8	100 vs 87.5	Diffuse dilation
Sigmoid volvulus	4 (11.1)	4/4	3/4	100 vs 75.0	Coffee-bean sign
Cecal volvulus	2 (5.6)	2/2	2/2	100 vs 100	Whirl sign
Overall	**36 (100)**	**34/36**	**29/36**	**94.4 vs 80.6***	-

*p=0.012, Multidisc: Multidisciplinary, Sens: Sensitivity, AI: Artificial intelligence (multidisc. AI vs ChatGPT-5.0)

Table 2. Guideline adherence and clinical outcomes

Parameter	Multidisciplinary AI	ChatGPT-5.0
Overall guideline adherence, n/N (%)	34/36 (94.4)	28/36 (77.8)*
Appendicitis: early surgery (WSES)	11/12 (91.7)	9/12 (75.0)
Cholecystitis: early cholecystectomy (Tokyo)	9/10 (90.0)	7/10 (70.0)
Volvulus: endoscopic/surgical (ASGE/WSES)	6/6 (100)	4/6 (66.7)
Ileus: conservative management (WJES)	8/8 (100)	7/8 (87.5)
ICU triage accuracy (SSC criteria)	33/36 (91.7)	28/36 (77.8)
Surgical vs conservative differentiation	27/28 (96.4)	23/28 (82.1)
Decision time, minutes (mean ± SD)	2.2±0.4	2.9±0.5
In-hospital mortality, n (%)	-	1 (2.8)

*p=0.018 vs. multidisciplinary, AI. WSES: World Society of Emergency Surgery, Tokyo: Tokyo Guidelines 2018, ASGE: American Society for Gastrointestinal Endoscopy, WJES: World Journal of Emergency Surgery, SSC: Surviving Sepsis Campaign, AI: Artificial intelligence, ICU:Intensive care unit, SD: Standard deviation

prospective head-to-head comparison with clinical teams is required. Krones et al. (10) demonstrated in Information Fusion that specialized model integration consistently outperforms single comprehensive systems in complex clinical scenarios requiring cross-domain expertise. Recent work by Wiest et al. (11) in Nature Reviews Gastroenterology & Hepatology highlighted that domain-specific large language models achieve optimal performance through integration rather than through general-purpose approaches. The 2025 multimodal AI review in World Journal of Gastroenterology emphasized the transformative potential of AI in precision diagnosis of gastrointestinal emergencies, with particular relevance to acute abdominal presentations. Our superior performance in differentiating appendicitis, cholecystitis, and ileus addresses a critical diagnostic challenge in which clinical and radiological overlap frequently delays appropriate management. This study has several limitations, including a small sample size, a single-center design, and retrospective data collection, all of which preclude definitive conclusions regarding generalizability; however, these preliminary results demonstrating high diagnostic accuracy within this limited pilot context support larger multicenter prospective validation trials to establish the clinical utility of multidisciplinary AI integration across the acute abdominal pain spectrum.

Ethics

Ethics Committee Approval: Approved by the Ankara Provincial Health Directorate Non-Interventional Clinical Research Ethics Committee (decision number: 2025-10-3, date: 24.10.2025).

Footnotes

Author Contributions

Concept: M.Ü., E.E., Design: M.Ü., E.E., Data Collection or Processing: M.Ü., Analysis or Interpretation: M.Ü., E.E., Literature Search: M.Ü., E.E., Writing: M.Ü., E.E.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Dadeh AA. Factors associated with unfavorable outcomes in patients with acute abdominal pain visiting the emergency department. BMC Emerg Med. 20226;22:195.
2. Yew KS, George MK, Allred HB. Acute abdominal pain in adults: evaluation and diagnosis. Am Fam Physician. 2023;107:585-96.
3. Wu YM, Tang FY, Qi ZX. Multimodal artificial intelligence technology in the precision diagnosis and treatment of gastroenterology and hepatology: innovative applications and challenges. World J Gastroenterol. 202514;31:109802.
4. Di Saverio S, Podda M, De Simone B, Ceresoli M, Augustin G, Gori A, et al. Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. World J Emerg Surg. 2020;15:27.
5. Yokoe M, Hata J, Takada T, Strasberg SM, Asbun HJ, Wakabayashi G, et al. Tokyo guidelines 2018: diagnostic criteria and severity grading of acute cholecystitis (with videos). J Hepatobiliary Pancreat Sci. 2018;25:41-54. Epub 2018 Jan 9.
6. Naveed M, Jamil LH, Fujii-Lau LL, Al-Haddad M, Buxbaum JL, Fishman DS, et al. American Society for Gastrointestinal Endoscopy guideline on the role of endoscopy in the management of acute colonic pseudo-obstruction and colonic volvulus. Gastrointest Endosc. 2020;91:228-235. Epub 2019 Nov 30. Erratum in: Gastrointest Endosc. 2020;91:721.
7. Dellinger RP, Rhodes A, Evans L, Alhazzani W, Beale R, Jaeschke R, et al. Surviving sepsis campaign. Crit Care Med. 2023;51:431-44.
8. Schouten D, Nicoletti G, Dille B, Chia C, Vendittelli P, Schuurmans M, et al. Navigating the landscape of multimodal AI in medicine: a scoping review on technical challenges and clinical applications. Med Image Anal. 2025;105:103621.
9. Kaczmarczyk R, Wilhelm TI, Martin R, Roos J. Evaluating multimodal AI in medical diagnostics. NPJ Digital Medicine. 2024;7:205.
10. Krones F, Marikkar U, Parsons G, Szmul A, Mahdi A. Review of multimodal machine learning approaches in healthcare. Information Fusion. 2025;114:102690.
11. Wiest IC, Bhat M, Clusmann J, Schneider CV, Jiang X, Kather JN. Large language models for clinical decision support in gastroenterology and hepatology. Nat Rev Gastroenterol Hepatol. 2025;22:773-87.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.18091
Eurasian J Emerg Med. 2026;25: 275-80

The Association of Systemic Inflammatory Indices with Alvarado Score-based Risk Stratification in Acute Appendicitis: A Retrospective Study

Serhat Örün[1], Abdullah Talha Demirel[1], Ufuk Çağman[1], Bayram Yıldırım[2], Murat Balkaya[1], Mehmet Çelik[3], Feyza Genç[1]

[1]Tekirdağ Namık Kemal University Faculty of Medicine, Department of Emergency Medicine, Tekirdağ, Türkiye
[2]Nusaybin State Hospital, Clinic of Emergency Medicine, Mardin, Türkiye
[3]Bandırma Training and Research Hospital, Clinic of Emergency Medicine, Balıkesir, Türkiye

Abstract

Aim: This study aimed to evaluate the association between systemic inflammatory indices, including the systemic immune-inflammation index (SII), neutrophil-to-lymphocyte ratio (NLR), and platelet-to-lymphocyte ratio (PLR), and the Alvarado score for risk stratification in patients presenting to the emergency department with a diagnosis of acute appendicitis (AA).

Materials and Methods: This retrospective descriptive study was approved by the Tekirdağ Namık Kemal University Faculty of Medicine Ethic Committee (approval number: 2025.70.04.06, date: 30.07.2025). A total of 69 adult patients who met inclusion criteria and presented with abdominal pain to the emergency department between January 1, 2023, and January 1, 2024 were enrolled. Among the patients, 53.6% were female and 46.4% male, with a mean age of 41.25±17.46 years. No significant age difference was observed between genders (p=0.8). The diagnosis of AA was confirmed by contrast-enhanced abdominal computed tomography in all cases. **Results:** ROC curve analysis was performed to assess the ability of inflammatory indices to discriminate among Alvarado score-defined risk categories. According to these analyses, NLR showed significant discriminatory performance across Alvarado score-defined high-risk categories with p=0.004 and [area under the curve (AUC)=0.81 95% confidence interval (CI): 0.71-0.91], and SII did so with p=0.009 and AUC=0.78 (95% CI: 0.66-0.90), whereas PLR was less predictive (p=0.16, AUC=0.65).

Conclusion: Our findings suggest that SII shows meaningful association with Alvarado score-based risk stratification and should not be interpreted as a standalone diagnostic marker. However, these markers should be used in conjunction with clinical assessment, and larger, multicenter prospective studies are needed to confirm their usefulness.

Keywords: Acute appendicitis, emergency department, neutrophil-to-lymphocyte ratio, systemic immune-inflammation index, platelet-to-lymphocyte ratio

Introduction

The appendix is a finger-shaped sac 5 to 10 cm long and 3 to 6 mm in diameter that extends from the cecum. Acute appendicitis (AA) is inflammation of the appendix, which is located in the abdominal cavity. AA is one of the most common surgical emergencies encountered among patients presenting to the emergency department. If it is not diagnosed promptly, it can lead to serious complications such as perforation, peritonitis, and sepsis. The lifetime risk of developing AA in the general population is reported to be approximately 7-8%, and the symptoms and clinical findings of the disease can vary from patient to patient. While AA usually presents with symptoms such as right lower quadrant pain, loss of appetite, nausea, and vomiting, these symptoms are non-specific,

Corresponding Author: Serhat Örün MD, Tekirdağ Namık Kemal University Faculty of Medicine, Department of Emergency Medicine, Tekirdağ, Türkiye
E-mail: serhatorun@gmail.com **ORCID ID:** orcid.org/0000-0001-5879-7858

Cite this article as: Örün S, Demirel AT, Çağman U, Yıldırım B, Balkaya M, Çelik M, et al. The association of systemic inflammatory indices with Alvarado score-based risk stratification in acute appendicitis: a retrospective study. Eurasian J Emerg Med. 2026;25: 275-80.

Received: 10.12.2025
Accepted: 29.03.2026
Published: 02.04.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

and a significant proportion of patients present with atypical features. This complicates diagnosis and can lead to unnecessary imaging tests or overlooking false-negative cases.

While physical examination and clinical evaluation are crucial in the diagnosis of AA, they alone do not provide sufficient diagnostic certainty. Laboratory tests [white blood cell (WBC) count, C-reactive protein (CRP) level, etc.] and imaging methods (ultrasound, computed tomography) are widely used to support the diagnosis. However, when access to imaging is limited or when radiation exposure is undesirable (e.g., in pregnant women), alternative diagnostic approaches are needed. In recent years, the potential role of inflammatory markers in the diagnosis of AA has been increasingly investigated.

Systemic inflammation indices are biochemical parameters that assess the inflammatory process in the body and are increasingly used in the diagnosis of acute and chronic inflammatory diseases. Among these parameters, the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and systemic immune-inflammation index (SII) stand out as promising biomarkers, particularly in the differential diagnosis of surgical and infectious diseases. Many studies have reported significant changes in NLR and PLR during acute inflammation, while SII has been reported to demonstrate stronger associations with inflammatory burden in diagnostic studies.

The Alvarado score, defined by Dr. Alvarado in 1986, is based on clinical and laboratory findings for the diagnosis of AA. Its purpose is to support the diagnosis and reduce unnecessary laparotomies (1). Importantly, the Alvarado score is not a gold standard diagnostic test but a structured risk stratification system intended to guide clinical decision-making. Studies have reported that the SII demonstrates a statistically significant correlation with the Alvarado score and has been reported to show high specificity and sensitivity in previous diagnostic studies (2).

This study was designed to explore the association between systemic inflammatory indices and Alvarado score-based risk stratification in patients with computed tomography-confirmed AA, rather than assess diagnostic accuracy against a non-appendicitis population. The findings are anticipated to expedite risk management in emergency departments, reduce unnecessary imaging studies, and improve patient management.

Materials and Methods

Our retrospective descriptive study received approval from This retrospective descriptive study was approved by the Tekirdağ Namık Kemal University Faculty of Medicine Ethic Committee (approval number: 2025.70.04.06, date: 30.07.2025). Patients presenting to the emergency department of Tekirdağ Namık Kemal University Faculty of Medicine Hospital with abdominal pain between January 1, 2023, and January 1, 2024, and diagnosed with AA on imaging were included in this study.

All study expenses were covered by the study directors; no financial assistance was used. Sixty-nine patients who presented to the emergency department with abdominal pain between the specified dates and who met the inclusion criteria were included. Inclusion criteria for the study were age 18-80, presentation to the emergency department between January 1, 2023, and January 1, 2024, and a diagnosis of AA following examinations in the emergency department. Exclusion criteria for the study were the following: being under 18 or over 80 years of age; lack of access to required study data; personal or family history of FMF; and refusal of treatment, leaving the emergency department, or referral from the emergency department after diagnosis. Patients with documented active systemic infection, inflammatory bowel disease, malignancy, or autoimmune inflammatory conditions were excluded if identifiable on review of medical records.

The electronic files of patients included in the study were reviewed, and their complaints, vital signs, physical examination findings, and laboratory findings at presentation were recorded on a precreated case report form. The neutrophil/lymphocyte ratio, platelet/lymphocyte ratio, systemic inflammation index, and Alvarado score were calculated. Platelet, neutrophil, and lymphocyte counts were recorded as cells/μL; therefore, SII values appear numerically higher than those reported in studies using $\times 10^9$/L units.

Statistical Analysis

All eligible cases within the specified dates were included in the retrospective analysis. Data were analyzed using SPSS version 18. Categorical variables were presented as counts and percentages and compared using chi-square tests. The Kolmogorov-Smirnov test assessed the normality of continuous variables. Normally distributed continuous variables were expressed as mean ± standard deviation and compared using Student's t-test, whereas non-normally distributed variables were presented as median (interquartile range) and compared using the Mann-Whitney U test. Fisher's exact test or chi-square test was applied for group comparisons, depending on the nature of the data. ROC curve analyses were conducted to evaluate the discriminatory ability of NLR, PLR, and SII across Alvarado score-defined risk categories. A p-value <0.05 was considered statistically significant.

Results

A total of 69 patients diagnosed with AA and admitted to the ward were included. Among them, 37 (53.6%) were female and 32

(46.4%) were male. The mean age was 41.25±17.46 years. There was no statistically significant age difference between females (40.78±16.73) and males (41.78±18.52) (p=0.8). The diagnosis of AA was confirmed by contrast-enhanced abdominal CT in all cases. All patients presented with abdominal pain. On abdominal examination, guarding and rebound tenderness were noted as overlapping clinical findings: guarding was present in 24 patients (34.8%), rebound tenderness was present in 23 patients (33.3%), and both signs were present in 12 patients (17.4%). Thirty-four patients (49.3%) had neither of these findings (Table 1).

Vital signs included a mean arterial pressure of 102.39 mmHg, an oxygen saturation of 98%, a temperature of 36.2 °C, and a pulse rate of 80 bpm (Table 2).

Laboratory parameters showed mean CRP 59.5±82.3 mg/L, WBC 13.6±11.5 $\times 10^3$/µL, neutrophils 9.4±3.6 $\times 10^3$/µL, lymphocytes 3.2±10.7 $\times 10^3$/µL, and platelets 278±87 $\times 10^3$/µL. The mean NLR was 5.89±5.71, PLR 162.27±103.24, and SII 11,131±1,543 (Table 3).

According to standard Alvarado score classification, patients were stratified into three risk groups: low risk (≤4), intermediate risk (5-6), and high risk (≥7). The mean Alvarado score was 6.04±1.89.

ROC analysis in the high risk (≥7) group showed NLR with p=0.004 and AUC=0.81 (0.71-0.91), SII with p=0.009 and AUC=0.78 (0.66-0.90), and PLR with p=0.16 and AUC=0.65 (0.49-0.81).

In the intermediate-risk (5-6) group, NLR had p=0.002 and AUC=0.74 (0.62-0.86); SII had p<0.001 and AUC=0.78 (0.67-0.89); and PLR had p=0.07 and AUC=0.63 (0.48-0.78).

In the low risk (≤4) group, NLR showed p=0.013 and AUC=0.32 (0.19-0.45), SII p<0.023 and AUC=0.33 (0.20-0.46), and PLR p=0.35 and AUC=0.43 (0.29-0.57).

For patients with Alvarado scores ≥5, ROC analyses revealed NLR with p=0.004 and AUC=0.74 (0.60-0.88), SII with p<0.001 and AUC=0.80 (0.65-0.94), and PLR with p=0.05 and AUC=0.67 (0.50-0.84).

In the low-risk group (Alvarado score ≤4), all inflammatory indices exhibited limited discriminatory performance (AUC <0.50).

Within the study population, NLR demonstrated its highest discriminatory performance at a cut-off value of 4.34, while PLR and SII showed optimal discrimination at cut-off values of 0.118 and 1.235, respectively. These cut-off points were associated with greater specificity in the higher Alvarado score categories, indicating improved discriminative capacity as clinical risk increased. Logistic regression analysis evaluating the association of Alvarado score, SII, NLR, and PLR with higher-risk categories showed coefficients for NLR=0.1639 (p=0.184), PLR=-0.0088 (p=0.065), and SII=0.0016 (p=0.008). The distribution of Alvarado score categories with NLR, PLR, and SII data is detailed in Figure 1.

Table 1. Demographic data and physical examination findings of the cases

Data	n (%), mean ± SD
Age	41.25±17.46
Gender Female Male	 37 (53.6%) 32 (46.4%)
Guarding Rebound tenderness Both signs Neither findings	24 (34.8%) 23 (33.3%) 12 (17.4%) 34 (49.3%)

Percentages are calculated for overlapping binary clinical findings. Categories are not mutually exclusive; therefore, percentages do not sum to 100%. SD: Standard deviation

Table 2. Vital signs of the cases

Vital signs	n (%), mean ± SD
Mean arterial pressure	102.39±16.32 mmHg
Pulse	80±16/minutes
Saturation	98±2%
Temperature	36.2±0.18°C

SD: Standard deviation

Table 3. Laboratory data of the cases

Data	mean ± SD
CRP	59.5±82.3
WBC	13.6±11.5
Neutrophil	9.4±3.6
Lymphocyte	3.2±10.7
Platelet	278±87
Neutrophil-to-lymphocyte ratio	5.89±5.71
Platelet-to-lymphocyte ratio	162.27±103.24
Systemic inflammation index	11.13 ±1.54

SD: Standard deviation CRP: C-reactive protein, WBC: White blood cell

Discussion

This study demonstrates that systemic inflammatory indices are associated with Alvarado score-based risk stratification and do not provide independent diagnostic accuracy for AA. By analyzing the contribution of SII to the diagnostic process, we aimed to obtain data to better understand the role of laboratory markers in diagnosing AA. We anticipate that our study findings may expedite the diagnostic process, reduce unnecessary imaging studies, and improve patient management. We believe, based on the literature, that stating that gender is not significant

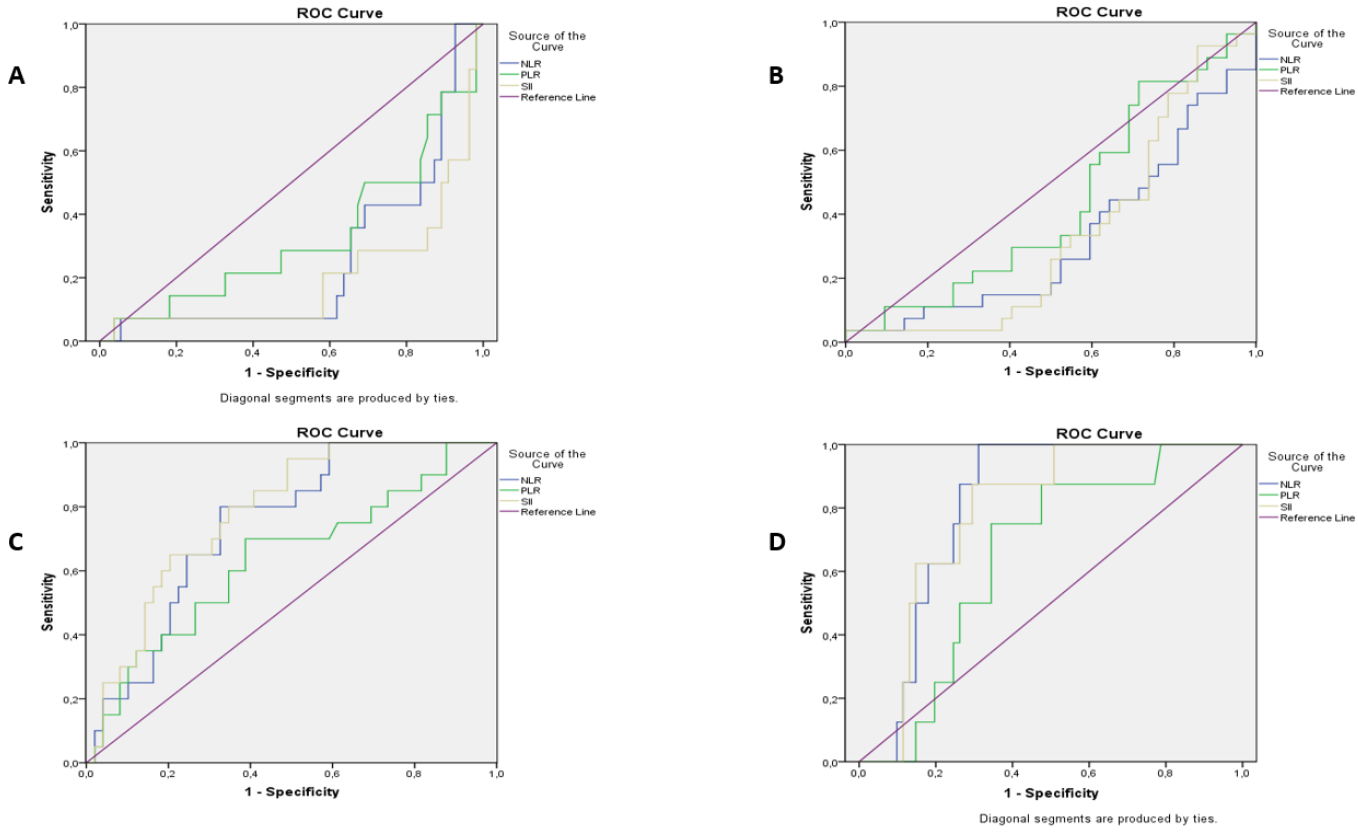


Figure 1. ROC analysis of groups with NLR, PLR and SII according to Alvarado score. A) The Alvarado score group is 4 or below (no risk), B) The Alvarado score group is 5-6 (low risk), C) The Alvarado score group is 7-8 (moderate risk), D) The Alvarado score group is 9-10 (high risk)
NLR: Neutrophil-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, SII: Systemic immune-inflammation index

for the diagnosis of AA would not be a misstatement. While the proportion of female (49.5%) and male (50.5%) patients was very similar in the study by Al Amri et al. (3), the study by Berhuni et al. (4) found 54.65% to be male, and similarly, the study by Shrestha et al. (5) found 54% to be male. Rajalingam et al. (6) reported 41.3% to be male. In our study, 46.4% of the patients were found to be male. Based on these data, a patient's gender alone does not appear to be a reliable determinant of AA diagnosis.

Regarding age distribution, AA typically affects young adults presenting with acute abdomen. Previous studies reported mean ages ranging from 28.58±16.65 to 38.16±14.50 (5-9). Our study found a mean age of 41.25±17.46, slightly higher but attributable to increasing life expectancy in the population.

The Alvarado score has been repeatedly evaluated to determine its effectiveness in preventing unnecessary imaging and laparotomy in the diagnosis of AA. We found a significant correlation between NLR and the Alvarado score. Hajibandeh et al. (10) reported NLR sensitivity of 90.91% and specificity of 88.89% for diagnosing and grading appendicitis severity. Kahramanca et al. (11) indicated that NLR aids in diagnosing and differentiating simple from complicated appendicitis, with 65.3% sensitivity and 54.7% specificity, complementing clinical examination. Ahmad et al. (12) reported sensitivity and specificity of 75.23% and 68.70%, respectively, for NLR in AA diagnosis. Markar et al. (13) ROC analysis yielded an AUC of 0.83 for NLR. Our results demonstrate statistically significant discrimination among low-, moderate-, and high-risk AA groups using NLR, suggesting its utility for clinical risk stratification and classification.

Rajalingam et al. (6) investigated PLR and NLR as surrogate biomarkers to differentiate complicated from uncomplicated appendicitis, finding both statistically significant, though PLR (sensitivity 75.9%, specificity 40.8%) lagged behind NLR (sensitivity 80.6%, specificity 47.2%). Kalayci and Kartal (14) found no significant difference in NLR and PLR between complicated and uncomplicated AA (p=0.926 and p=0.642, respectively). Chen et al. (15) compared NLR and PLR efficacy for differentiating AA from right ureterolithiasis, concluding that NLR (specificity 81.82%, sensitivity 68.87%) was more accurate and reliable than PLR (specificity 84.85%, sensitivity 37.74%). Celik et al. (16) reported

that both NLR (61.1% sensitivity, 73.2% specificity) and PLR (42% sensitivity, 86% specificity) effectively differentiate complicated from uncomplicated AA, with NLR having superior sensitivity and specificity. Our data indicate that, although PLR has lower discriminatory capacity than NLR, it can serve as a supportive biomarker above certain threshold values.

Şener et al. (17) reported that SII sensitivity for AA diagnosis was 82% and specificity 66.7%, also playing a role in distinguishing complicated and uncomplicated appendicitis. Mutlu et al. (18) identified SII as an independent predictor for complicated AA and noted its superior predictive power (AUC 0.809) over NLR (AUC 0.729) and CRP (AUC 0.732). Aydın and Tatlıparmak (19) found that SII had 78% sensitivity and 79% specificity for the diagnosis of AA (AUC 0.81). Our findings align with the literature, supporting that SII showed the strongest associations with higher clinical risk categories, reflecting their relationship with the magnitude of systemic inflammation.

Reported sensitivity and specificity of the Alvarado score at cut offs ≥5 and ≥7 vary across studies. Awayshih et al. (20) concluded that the Alvarado score is not sufficiently sensitive, reporting 54% sensitivity and 75% specificity. Din et al. (21) prospective study demonstrated that a cutoff value of 7.0 yielded 71.1% sensitivity and 75.8% specificity. Bouali et al. (22) reported higher diagnostic performance at increased Alvarado scores, with a decline in positive predictive value as the score decreased. In line with these findings, our results indicate that higher Alvarado score categories are associated with improved discriminatory performance of SII, NLR, and PLR across risk strata, with SII demonstrating the strongest association with the scoring system.

Study Limitations

• The absence of a non-appendicitis control group precludes the assessment of true diagnostic accuracy metrics, such as sensitivity and specificity.

• Single-center data collection limits generalizability; multicenter studies with larger samples are recommended.

• The sample size was limited. Larger cohorts would allow advanced analyses, including subgroup analyses.

• A retrospective design limits causal inference; prospective studies would better evaluate the efficacy of clinical decision-making.

• No distinction between complicated and uncomplicated appendicitis was made; this differentiation could strengthen the results, given the higher inflammatory markers observed in complicated cases.

• Potential inaccuracies or missing data inherent in retrospective electronic medical records may have influenced the results.

Conclusion

Our study suggests that SII is a supportive biomarker, parallel to the Alvarado score, in assessing the risk of AA. Both SII and NLR can be used in conjunction with the Alvarado score for risk assessment. However, these markers alone are insufficient, and larger-scale, multicenter, prospective studies are needed to improve diagnostic accuracy.

Ethics

Ethics Committee Approval: The study protocol was approved by the Ethics Committee of Tekirdağ Namık Kemal University Faculty of Medicine (approval number: 2025.70.04.06, date: 30.07.2025).

Informed Consent: This is a retrospective descriptive study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: S.Ö., A.T.D., U.Ç., B.Y., M.B., M.Ç., Concept: S.Ö., A.T.D., U.Ç., B.Y., F.G., Design: S.Ö., A.T.D., U.Ç., B.Y., F.G., Data Collection or Processing: S.Ö., A.T.D., U.Ç., B.Y., M.B., M.Ç., F.G., Analysis or Interpretation: S.Ö., Literature Search: S.Ö., A.T.D., U.Ç., B.Y., M.B., M.Ç., F.G., Writing: S.Ö., A.T.D., U.Ç., B.Y.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med. 1986;15:557-64.
2. Arredondo Montero J, Delgado-Miguel C, Pérez Riveros BP, Fernández-Atuan R, Rico-Jiménez M. Systemic immune-inflammation index for the diagnosis of acute appendicitis: a systematic review. J Surg Res. 2025;309:88-102.
3. Al Amri FS, Fihrah RS, Al Jabbar I, Alqahtani R, Alnujaymi B, Alshehri RM, et al. Accuracy of neutrophil-to-lymphocyte ratio in predicting the severity of acute appendicitis: a single-center retrospective study. Cureus. 2023;15:e45923.
4. Berhuni MS, Yönder H, Elkan H, Koyuncu MH, Özgül F, Tatlı F, et al. Diagnostic value of systemic immune inflammation index in identifying complicated acute appendicitis cases. Cureus. 2024;16:e73046.
5. Shrestha B, Koju R, Makaju Shrestha S, Shrestha K, Hada G, Shakya YR, et al. Predicting complicated appendicitis based on neutrophil to lymphocyte ratio, a study in University Hospital of Nepal. Kathmandu Univ Med J (KUMJ). 2023;21:270-4.
6. Rajalingam VR, Mustafa A, Ayeni A, Mahmood F, Shammout S, Singhal S, et al. The role of neutrophil-lymphocyte-ratio (NLR) and platelet-lymphocyte-ratio (PLR) as a biomarker for distinguishing between complicated and uncomplicated appendicitis. Cureus. 2022;14:e21446.

7. Halaseh SA, Kostalas M, Kopec CA, Nimer A. Single-center retrospective analysis of neutrophil, monocyte, and platelet to lymphocyte ratios as predictors of complicated appendicitis. Cureus. 2022;14:e29177.
8. Senol S, Kusak M, Özdemir DB, Sendil AM. Diagnostic value of serum sodium level and neutrophil-to-lymphocyte ratio in predicting severity of acute appendicitis: a retrospective cross-sectional two-center study. Medicina (Kaunas). 2024;60:1844.
9. Białas M, Taran K, Gryszkiewicz M, Modzelewski B. Wartość wskaźnika neutrofilowo-limfocytarnego w diagnostyce zapalenia wyrostka robaczkowego [Evaluation of neutrophil-lymphocyte ratio usefulness in the diagnosis of appendicitis]. Wiad Lek. 2006;59:601-6.
10. Hajibandeh S, Hajibandeh S, Hobbs N, Mansour M. Neutrophil-to-lymphocyte ratio predicts acute appendicitis and distinguishes between complicated and uncomplicated appendicitis: a systematic review and meta-analysis. Am J Surg. 2020;219:154-63. Epub 2019 Apr 27.
11. Kahramanca S, Ozgehan G, Seker D, Gökce EI, Seker G, Tunç G, et al. Neutrophil-to-lymphocyte ratio as a predictor of acute appendicitis. Ulus Travma Acil Cerrahi Derg. 2014;20:19-22.
12. Ahmad KA, Ideris N, Aziz SHSA. A cross-sectional study of neutrophil-to-lymphocyte ratio in diagnosing acute appendicitis in Hospital Melaka. Malays J Med Sci. 2019;26:55-66.
13. Markar SR, Karthikesalingam A, Falzon A, Kan Y. The diagnostic value of neutrophil: lymphocyte ratio in adults with suspected acute appendicitis. Acta Chir Belg. 2010;110:543-7.
14. Kalayci T, Kartal M. Significance of neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio, serum albumin and prognostic nutritional index as predictors of morbidity in super-elderly patients operated on for acute appendicitis. Eur Rev Med Pharmacol Sci. 2022;26:820-7.
15. Chen L, Zhang Y. Comparison of the diagnostic values of leukocytes, neutrophils, neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio in distinguishing between acute appendicitis and right ureterolithiasis. Clin Lab. 2020;66.
16. Celik B, Nalcacioglu H, Ozcatal M, Altuner Torun Y. Role of neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio in identifying complicated appendicitis in the pediatric emergency department. Ulus Travma Acil Cerrahi Derg. 2019;25:222-8.
17. Şener K, Çakır A, Kılavuz H, Altuğ E, Güven R. Diagnostic value of systemic immune inflammation index in acute appendicitis. Rev Assoc Med Bras (1992). 2023;69:291-6.
18. Mutlu H, Sert ET, Kokulu K, Uslu Y. Systemic immune inflammation index may be a new powerful marker for the accurate early prediction of complications in patients with acute appendicitis. Cukurova Medical Journal. 2023;48:679-85.
19. Aydın ÖF, Tatlıparmak AC. Diagnostic accuracy of systemic immune-inflammation index for acute appendicitis in the geriatric population. Eur J Med Res. 2025;30:63.
20. Awayshih MMA, Nofal MN, Yousef AJ. Evaluation of Alvarado score in diagnosing acute appendicitis. Pan Afr Med J. 2019;34:15.
21. Din SSU, Ullah Baig I, Tassawar Hussain M, Sadiq A, Humayun T, et al. RIPASA versus Alvarado score in the assessment of acute appendicitis: a prospective study. Turk J Surg. 2023;39:231-6.
22. Bouali M, El Berni Y, Moufakkir A, El Bakouri A, El Hattabi K, Bensardi F, et al. Value of Alvarado scoring system in diagnosis of acute appendicitis. Ann Med Surg (Lond). 2022;77:103642.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Review
DOI: 10.4274/eajem.galenos.2026.89814
Eurasian J Emerg Med. 2026;25: 281-7

Aeromedical Patient Transfer in a High-volume Health System: Centralised Coordination and Governance in Türkiye

Behçet Al[1], Yunus Emre Gemici[2]

[1]Göztepe Prof. Dr. Süleyman Yalçın City Hospital, İstanbul Medeniyet University, Faculty of Medicine, Department of Emergency Medicine, İstanbul, Türkiye
[2]Göztepe Prof. Dr. Süleyman Yalçın City Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye

Abstract

Aeromedical patient transfer plays a crucial role in ensuring continuity of time-sensitive care in health systems characterized by high patient volumes, geographic diversity, and regional disparities in access to healthcare. This review examines the core organizational and governance principles of aeromedical patient transfer systems and evaluates the centrally coordinated aeromedical model implemented in Türkiye, in the context of the global literature, to generate context-sensitive policy insights for high-volume health systems.

A narrative, policy-informed review approach was adopted. The global literature on aeromedical patient transfer was examined, with a focus on organizational models, central coordination mechanisms, clinical prioritization, medical leadership, and safety-oriented governance. Türkiye's air ambulance system was analyzed against national regulatory frameworks and operational structures and contextualized through comparison with international practices.

Türkiye has developed a centrally coordinated aeromedical model integrating air ambulance services into the national emergency medical services system. The model is characterized by centralized decision-making, standardized clinical indications, strong medical leadership, and systematic integration of operational safety and risk management. Aeromedical resources are treated as limited, high-value health assets, with deployment guided by clinical benefit and system-level priorities rather than transport speed alone.

The Turkish experience suggests that the effectiveness of aeromedical patient transfer in high-volume health systems depends more on governance, coordination, and rational resource allocation than on fleet size or technological capacity. While not intended for direct replication, this centrally coordinated model offers adaptable principles that may support coordinated access to specialized care in other middle-income countries

Keywords: Aeromedical transfer, coordination, health system, Türkiye

Introduction

Aeromedical patient transfer should be seen not just as a way to physically move patients from one place to another, but as a key part of the health system focused on reducing geographic and structural disparities in access to healthcare. Especially for populations in rural and remote areas, air ambulance systems offer a complementary service that allows timely access to definitive care for time-critical clinical conditions. Increasingly, it is emphasized that the effect of aeromedical transfers on health outcomes depends less on flight capacity or fleet size and more on how these services are organized, coordinated, and governed within the health system (1,2).

Aeromedical services originated from military needs, particularly during the First and Second World Wars, when they were developed to facilitate the rapid evacuation of wounded soldiers to medical facilities. Over time, these services were integrated into civilian health systems and became widely used to manage trauma, acute cardiovascular events, stroke, obstetric emergencies, and intensive care patients (3,4). Today, aeromedical transfers conducted by fixed-wing and rotary-wing aircraft constitute a key component of emergency medical services in many countries. However, substantial variation exists across countries in organizational structures, utilization indicators, and governance models for aeromedical services (2).

Corresponding Author: Prof. MD. Behçet Al, Göztepe Prof. Dr. Süleyman Yalçın City Hospital, İstanbul Medeniyet University, Faculty of Medicine, Department of Emergency Medicine, İstanbul, Türkiye
E-mail: behcetal@gmail.com **ORCID ID:** orcid.org/0000-0001-8743-8731

Cite this article as: Al B, Gemici YE. Aeromedical patient transfer in a high-volume health system: centralized coordination and governance in Türkiye. Eurasian J Emerg Med. 2026;25: 281-7.

Received: 20.01.2026
Accepted: 31.03.2026
Published: 06.04.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Globally, the configuration of aeromedical systems varies considerably across countries, depending on their geographic characteristics, distribution of health infrastructure, economic capacity, and governance approaches. In countries with large land areas, high proportions of rural populations, or limited access to healthcare facilities, aeromedical patient transfer is positioned as a core service delivery modality. In contrast, in countries with well-developed road networks and extensive emergency medical services infrastructure, aeromedical transport tends to be used more selectively and based on stricter indications (2,5). This diversity underscores the need to examine aeromedical transfers not only within a clinical framework but also in relation to health system design and policy choices.

Türkiye is a middle-income country with high patient volumes, serving a population exceeding 86 million and recording more than one billion healthcare encounters annually. Meeting healthcare demand at this scale particularly for time-sensitive conditions and cases requiring advanced levels of care is not sustainable without effective referral pathways and robust coordination mechanisms (6). The country's geographic diversity, disparities in rural-urban healthcare distribution, and the concentration of specialized centers in specific regions have rendered aeromedical patient transfer a strategic, rather than merely complementary, instrument within the Turkish health system (7).

This review aims to briefly outline the global development and core organizational principles of aeromedical patient transfer systems and to examine the centrally coordinated air ambulance governance model implemented in Türkiye, a middle-income country with a high-volume health system. The review adopts a comprehensive approach to analyzing the structural framework, governance, and decision-making processes, team composition, clinical indications, resource prioritization, and safety and risk management mechanisms of Türkiye's aeromedical system. Drawing on the Turkish experience, the study seeks to present adaptable policy lessons for global health systems on how to organize aeromedical services at the national level in a rational, safe, and sustainable manner (Table 1).

Materials and Methods

This study is a narrative review that adopts a policy-informed health systems analysis approach. The review summarizes the core organizational framework of the global literature on aeromedical patient transfer and focuses analytically on the organizational structure, team composition, and coordination mechanisms of the centrally coordinated air ambulance system implemented in Türkiye.

The literature search was conducted to include publications addressing the organizational, governance, and coordination roles of aeromedical services within health systems. Key concepts,

Table 1. Comparison of aeromedical system models: international patterns and the centralized model in Türkiye

Domain	Common International/fragmented models	Centralized Model in Türkiye
Coordination structure	Aeromedical transfers are often coordinated by regional, local, or multiple dispatch centers	Transfer requests are evaluated through a centrally operated national coordination structure
Dispatch decision-making	Decisions may vary across regions depending on local practice, available resources, and institutional preferences	Decisions are made through a centralized mechanism using nationally standardized criteria
Resource allocation	Allocation may differ between regions and may be influenced by local capacity and operational variability	Aeromedical resources are managed as nationally coordinated, limited, high-value assets
Clinical indication control	Indications may be broader or less standardized across systems	Transfers are prioritized according to clinical benefit, time sensitivity, and system-level need
Risk of over-triage/ inappropriate use	Systems lacking central prioritization may be more vulnerable to overuse or inefficient deployment	Central filtering aims to reduce inappropriate utilization and preserve resource efficiency
Medical leadership	Team composition and physician involvement vary across countries and services	Aeromedical transfers are conducted by physician-led teams supported by paramedics
Integration with the health system	Integration with emergency medical services and referral systems may be partial or variable	Air ambulance services are integrated into the national emergency medical services system.
Governance model	Governance structures are often heterogeneous and may involve regional or mixed administrative approaches	Governance is centrally coordinated under the Ministry of Health with standardized oversight
Safety and risk management	Safety practices vary depending on local operational structures and aviation oversight	Dispatch decisions explicitly incorporate both clinical appropriateness and aviation safety considerations
Operational logic	Aeromedical use may be influenced primarily by transport speed or regional accessibility	Deployment is guided by clinical appropriateness, operational safety, and rational resource prioritization rather than speed alone

EMS: Emergency medical services

including aeromedical patient transfer, aeromedical services, patient transfer models, organization, coordination, and clinical indications, were utilized individually and in combination. This approach aimed to prioritize studies reflecting system- and policy-level dimensions, rather than detailed clinical or technical aspects.

The literature search was conducted using a concept-based strategy focusing on system-level and policy-oriented aspects of aeromedical patient transfer. Rather than applying a formal systematic search protocol, the review prioritized studies addressing organizational structure, governance models, coordination mechanisms, and resource allocation. This approach was chosen to align with the review's narrative and policy-informed nature. No formal language restrictions were applied during the literature selection process. Official documents and national regulatory sources were identified through institutional databases and government publications relevant to the Turkish aeromedical system.

Priority was given to studies examining the historical development of aeromedical transfers, organizational models used across countries, and the role of air ambulance services in high-volume health systems. In selecting the literature, particular emphasis was placed on recent publications that provide system-level insights into health system design, governance, resource utilization, and coordination mechanisms. Clinical case series, studies focusing primarily on technical flight characteristics, and publications lacking organizational or managerial evaluation were excluded from the review.

Country-specific analyses for Türkiye were based on the Regulation on the Operation of Air Ambulance Aircraft issued by the Ministry of Health of the Republic of Türkiye, as well as the national air ambulance implementation framework developed in accordance with this regulation. This national framework was examined through a comparative lens alongside aeromedical transfer models described in the international literature, and similarities, differences, and context-specific characteristics were analyzed methodologically.

Findings derived from the analytical process were synthesized under the following thematic domains:

1. The global development and organizational models of aeromedical transfer,
2. The structural and governance characteristics of Türkiye's air ambulance system, and
3. The effects of organization and coordination on patient safety, resource utilization, and service sustainability.

Within this framework, the review assessed whether the centralized coordination model implemented in Türkiye offers adaptable policy implications for countries with comparable socioeconomic and geographic characteristics.

Türkiye's Centralized Aeromedical Coordination Model in a High-volume Health System

In Türkiye, aeromedical patient transfer has been structured not as a service model driven by individual requests or regional initiatives, but as a strategic component integrated into the national health system's central planning and governance. This approach positions aeromedical services not merely as a technical transport capacity, but as a managerial instrument that ensures continuity of time-sensitive care within a high-volume health system.

Under the coordination of the Ministry of Health of the Republic of Türkiye, air ambulance services are centrally dispatched and managed, with clinical appropriateness, operational safety, and resource prioritization evaluated in accordance with nationally standardized criteria (7). This structure aims to ensure that aeromedical resources are utilized rationally and transparently, even under conditions of high demand.

Structural and Governance Framework

The legal and administrative foundation of Türkiye's air ambulance system is defined by national regulations governing the operation of ambulance helicopters and fixed-wing ambulance aircraft. While the Ministry of Health assumes medical and managerial responsibility for the service, the operational management of aircraft is conducted through a public-private partnership model. This hybrid structure preserves the regulatory and supervisory role of public authority while supporting operational capacity through private-sector resources.

Flight safety and technical oversight processes are carried out by the Directorate General of Civil Aviation, thereby establishing an institutional division of responsibility between medical requirements and aviation safety (7). This dual-layered governance model enables the simultaneous safeguarding of patient safety and flight safety within aeromedical services.

Centralized Dispatch and Decision-making Mechanism

In Türkiye, aeromedical transfer requests are assessed by a centrally operated coordination unit within the Ministry of Health that operates 24/7 (7). This unit consolidates requests from across the country into a single operational pool, coordinates dispatch decisions, and prioritizes cases when multiple concurrent requests arise.

During the decision-making process, multiple factors are considered, including the patient's clinical condition, time sensitivity of the transfer, availability of ground ambulance alternatives, geographic distance, meteorological conditions, and flight safety considerations. This multidimensional approach aims to ensure that air ambulances are deployed not solely for speed advantage, but in situations where they provide a clear clinical and system-level benefit (8).

The centralized dispatch mechanism serves as a filter that limits inappropriate or excessive use of aeromedical resources, particularly during periods of high demand. International literature indicates that in systems lacking central prioritization mechanisms, air ambulances are more vulnerable to over-triage and inefficient resource use (9). The Turkish model offers a structural response to these risks.

Clinical Indications and Resource Prioritization

In Türkiye, aeromedical patient transfers are considered a scarce and high-cost healthcare resource and are prioritized based on clinical benefit. Time-sensitive trauma, acute cardiovascular and neurological emergencies, and patients requiring advanced intensive care constitute the primary clinical scenarios in which air ambulance utilization is emphasized.

Conversely, the restriction of aeromedical use in cases where ground transfer can be performed within a comparable timeframe and with adequate safety has been adopted as a fundamental principle. This approach seeks to preserve aeromedical services as a "rare and high-value" resource and to maximize overall system-level benefit. In this respect, the Turkish practice aligns with rational utilization principles recommended in the literature (8,10).

Team Structure and Medical Leadership

Aeromedical transfers in Türkiye are conducted by physician-led teams. While the medical team assumes clinical responsibility throughout the transfer process, the flight crew is responsible for the safe operation of the aircraft. This role distribution enables independent yet coordinated management of medical decision-making processes and aviation safety.

Medical leadership encompasses not only in-flight care but also pre-transfer assessment and determination of transfer appropriateness. This structure enables aeromedical transfers to be conceptualized as a process of mobile advanced care, rather than merely a logistical transport solution. The literature reports more rational utilization of aeromedical resources in systems characterized by strong physician leadership (1).

Safety, Risk Management, and Decision-Support Approach

Aeromedical patient transfers are high-risk operations requiring a continuous balance between patient benefit and flight safety. In Türkiye, dispatch and tasking decisions take into account not only clinical requirements but also aviation safety factors such as weather conditions, visibility, and crew availability.

Deferral or cancellation of transfers that are clinically appropriate but operationally unsafe is considered a natural outcome of a risk-based decision-making approach (10). The centralized operational structure also facilitates systematic monitoring of incidents and near-miss events, enabling their transformation into institutional learning processes.

System-level Evaluation

The Turkish experience demonstrates that the success of aeromedical patient transfer depends less on flight capacity than on centralized coordination, standardized clinical indications, medical leadership, and safety-oriented governance components. This model provides a context-sensitive example of how limited aeromedical resources can be managed equitably, safely, and sustainably within high-volume health systems.

Operational Capacity and Human Resources Structure

Air ambulance services in Türkiye are structured according to an operational capacity aligned with the centralized coordination model. Within the system operated by the Ministry of Health, 17 helicopters and 2 fixed-wing ambulance aircraft are currently in service (6,7). To ensure optimal benefits in terms of geographic characteristics and service accessibility, the country is divided into 17 operational regions, each operating under national central coordination (7). All aeromedical transfer requests are submitted to the central operations unit, and dispatch decisions are implemented following central approval.

Air ambulance aircraft are operated under a leasing model, whereas physicians and paramedics assigned to aeromedical missions are employed by the Ministry of Health (7). This structure enhances operational flexibility while ensuring that medical responsibility and clinical oversight remain under public authority. The joint deployment of physicians and paramedics within aeromedical teams supports the delivery of care according to a mobile advanced life support approach. In addition, healthcare personnel assigned to these aircraft are required to hold internationally recognized certifications, such as Advanced Trauma Life Support and Advanced Cardiac Life Support, thereby helping ensure standards of clinical competence and patient safety (7).

Discussion

This review examined the Turkish experience of aeromedical patient transfer in a high-volume health system, focusing on organizational structure, governance, clinical decision-making, and safety. The findings indicate that the effectiveness of aeromedical services is closely associated not only with flight capacity or technological infrastructure, but also with centralized decision-making, standardized clinical indications, and safety-oriented governance components (1,2). These observations are consistent with findings from the international literature, which emphasize the importance of coordinated dispatch systems, appropriate patient selection, and governance structures in optimizing aeromedical resource utilization.

The Role of Centralized Coordination in the Literature

International literature indicates that a substantial proportion of aeromedical systems are organized through regional, fragmented, or locally driven initiatives. In such structures, dispatch decisions made by multiple centers can complicate the standardization of clinical indications and lead to resource concentration in specific regions (2). In particular, systems lacking central prioritization mechanisms have been reported to carry an increased risk of overuse or inappropriate utilization of air ambulances (9).

The centralized coordination model implemented in Türkiye diverges from these fragmented structures by enabling national-level management of aeromedical resources through a single operational center. This approach represents a structural response to the problems of over-triage and resource waste that are frequently highlighted in the literature (8).

Clinical Indications and Resource Prioritization

In many countries, air ambulance utilization is applied across a broad range of indications on the assumption of time savings. However, studies published in recent years have demonstrated that air ambulances do not consistently confer the expected patient benefit across all clinical scenarios, underscoring the critical importance of appropriate patient selection (5). In the Turkish model, the centralized evaluation of aeromedical transfers based on clinical indications aligns with these findings.

In practice, limiting air ambulance use for patients who can be safely transferred by ground transport within a comparable timeframe reflects treating aeromedical resources as a “rare and high-value” service. This principle is consistent with rational utilization strategies recommended in the literature (8).

Physician Leadership and Medical Decision-making Processes

The perspective that aeromedical transfers constitute a process of mobile advanced care rather than a purely logistical activity has gained increasing prominence. Physician-led team models have been reported to provide important advantages in managing clinical risk and improving the quality of dispatch decisions (1). In Türkiye, the explicit definition of medical leadership in transfer processes is a key strength, aligning closely with international evidence.

Safety and Risk Management: Balancing Clinical Benefit and Flight Safety

Air ambulance operations inherently require a delicate balance between patient benefit and flight safety. International studies have reported higher rates of serious incidents and accidents in systems where safety-oriented risk assessment mechanisms are insufficiently structured (10). In Türkiye, the postponement or cancellation of clinically appropriate transfers when flight safety conditions are not acceptable reflects the adoption of a risk-based decision-making approach.

This approach is consistent with the literature emphasizing that aeromedical services should not operate under an “always fly” paradigm, but rather in accordance with principles of system safety and long-term sustainability (10).

System Effects of the Hybrid (Public-Private) Model

The delivery of air ambulance services through a public-private partnership model in Türkiye enables the expansion of operational capacity while maintaining medical and managerial control under public authority. Although similar hybrid models exist in several Organization for Economic Co-operation and Development countries, the literature suggests that dispatch mechanisms as centralized and standardized as those implemented in Türkiye remain relatively uncommon (2). This feature distinguishes the Turkish model from an organizational perspective.

Implications for Global Health Systems

The Turkish experience offers important lessons regarding the organization of aeromedical services in middle-income countries with high patient volumes. Centralized coordination, standardized clinical indications, physician leadership, and safety-oriented risk management may be regarded as core principles adaptable across diverse health system contexts. However, rather than direct replication, adapting these principles to countries’ specific geographic, economic, and governance conditions represents a more realistic and effective approach (8).

Study Limitations

This review has several limitations that should be acknowledged. First, the number of peer-reviewed scientific studies focusing specifically on aeromedical patient transfer in Türkiye remains limited. This reflects the relative scarcity of system-level academic analyses rather than deficiencies in operational capacity or institutional development. Consequently, several aspects of the Turkish aeromedical system are informed by grey literature and official policy documents.

Second, this review examines the experience of a single country. The findings are therefore not intended to be directly generalizable. Instead, the organizational structures and governance principles discussed should be interpreted as context-specific approaches that may offer adaptable insights for health systems operating under similar demographic, geographic, and service-volume conditions.

Third, the review adopts a narrative, organizational, and policy-oriented perspective, with a deliberate focus on governance, coordination, and decision-making mechanisms. As such, it does not provide a quantitative evaluation of patient-level outcomes, cost-effectiveness, or comparative clinical effectiveness of aeromedical versus ground transport modalities.

Finally, Türkiye's geographic diversity, population distribution, nationally integrated ground emergency medical services, and frequent exposure to large-scale disasters represent important contextual factors shaping the organization and utilization of aeromedical services. These system-level characteristics may influence both the necessity and functioning of the centralized coordination model described and should be considered when interpreting the broader policy implications.

Conclusion

The effective delivery of aeromedical patient transfer services in high-volume health systems appears to depend not only on aircraft availability or technological capacity, but also on how these services are organized, governed, and integrated into the broader health system. This review demonstrates that the aeromedical system in Türkiye is structured around a centralized coordination model that positions air ambulance services as a strategic component of the national health system, rather than a standalone transport modality.

The Turkish experience highlights the importance of centralized decision-making, standardized clinical indications, strong medical leadership within aeromedical teams, and the systematic incorporation of safety-oriented risk management into operational processes. Together, these elements support the rational use of limited and high-cost aeromedical resources, particularly in settings characterized by high patient volumes, geographic diversity, and time-sensitive care needs.

From a global health perspective, the Turkish model should not be viewed as a template for direct replication. Instead, it offers a set of adaptable organizational and governance principles that may inform the design of aeromedical services in other middle-income countries facing similar systemic pressures. When appropriately contextualized, aeromedical patient transfer can function not merely as an emergency transport solution but also as a policy instrument that contributes to health system resilience, equity, and coordinated access to specialized care.

Ethics

Ethics Committee Approval: This study is a narrative review based on previously published literature and publicly available data. Therefore, ethical approval and informed consent were not required.

Footnotes

Author Contributions

Concept: B.A., Y.E., Design: B.A., Data Collection or Processing: B.A., Y.E., Analysis or Interpretation: B.A., Literature Search: B.A., Writing: B.A., Y.E.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Low A, Hulme, J. ABC of transfer and retrieval medicine. Wiley-Blackwell. 2014.
2. Muyambi K, Gardiner F, Sollid S, Hyldmo PK, Yisma E, Spring B, et al. Aeromedical retrieval services characteristics globally: a scoping review. Scand J Trauma Resusc Emerg Med. 2022;30:71.
3. Posselt BN, Greenhalgh AM, Almond MK. 100 years of the Royal Air Force's contribution to medicine: providing care in the air and delivering care by air. Clin Med (Lond). 2018;18:297-300.
4. Barr J, Montgomery S. Helicopter medical evacuation in the Korean War: did it matter? J Trauma Acute Care Surg. 2019;87(1S Suppl 1):S10-S13.
5. Franklin RC, King JC, Aitken PJ, Elcock MS, Lawton L, Robertson A, et al. Aeromedical retrievals in Queensland: a five-year review. Emerg Med Australas. 2021;33:34-44.
6. Ministry of Health of the Republic of Türkiye. Health Statistics Yearbook 2023. Available from: https://ohsad.org/wp-content/uploads/2025/02/Saglik-Istatistikleri-Yilligi-2023-Turkce.pdf
7. Ministry of Health of the Republic of Türkiye. Organization and coordination of air ambulance services in Türkiye. Available from: https://saglik.gov.tr/TR-11147/ambulans-hava-araci-isletmesine-dair-esaslar.html
8. Lyng JW, Braithwaite S, Abraham H, Brent CM, Meurer DA, Torres A, et al. Appropriate air medical services utilization and recommendations for

integration of air medical services resources into the EMS system of care: a joint position statement and resource document of NAEMSP, ACEP, and AMPA. Prehosp Emerg Care. 2021;25:854-73.

9. Madiraju SK, Catino J, Kokaram C, Genuit T, Bukur M. In by helicopter out by cab: the financial cost of aeromedical over triage of trauma patients. J Surg Res. 2017;218:261-70.

10. Aherne BB, Zhang C, Chen WS, Newman DG. Preflight risk assessment for improved safety in helicopter emergency medical service operations. Aerosp Med Hum Perform. 2019;90:792-9.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.34467
Eurasian J Emerg Med. 2026;25: 288-93

A Retrospective Study of Prognostic Indicators of Mortality in In-hospital Cardiac Arrest in a Tertiary Care Emergency Center

Libania Angelina John, Viji Thamby Nalatha, Gideon Kevin Devasir, Priya Ganesan

Christian Medical College Vellore, Department of Emergency Medicine, Vellore, India

Abstract

Aim: Cardiac arrests are a critical public health issue, especially in developing countries like India, where access to timely advanced medical care is limited. This study aims to determine the predictors of mortality among patients sustaining in-hospital cardiac arrest (IHCA).

Materials and Methods: This retrospective observational cohort study in a tertiary hospital emergency department analyzed one-year IHCA patients. Demographics, causes, and outcomes were recorded from institutional database. Data entered in Excel were analyzed using SPSS version 25.0. Multivariate logistic regression was used to identify independent predictors of mortality among patients presenting with IHCA.

Results: Among 267 IHCA patients (mean age 52±17 years, male preponderance), hypoxia (43.44%) and coronary thrombosis (25.84%) were the most common reversible causes. Return of spontaneous circulation (ROSC) was achieved in 214 (80.14%) patients, while only 22 (8.23%) had favorable neurological outcomes (mRS <3) at hospital discharge. Logistic regression identified inability to attain ROSC and prolonged resuscitation (>20 minutes) as independent predictors of mortality.

Conclusion: Our study underscores the critical importance of achieving ROSC within 20 minutes, through early management of reversible causes, and consistent high-quality chest compressions, which are pivotal for favorable neurological outcomes among survivors.

Keywords: In-hospital cardiac arrest, mortality predictors, return of spontaneous circulation, emergency department, cardiopulmonary resuscitation

Introduction

Cardiac arrest occurs when the heart stops resulting in a sudden cessation of blood circulation throughout the body. This catastrophic event deprives vital organs, most critically the brain, of oxygen, causing rapid cell death if not immediately addressed (1,2).

Cardiac arrest represents a significant public health concern due to its high incidence and mortality rates (3,4). In India, recent reports suggest that in-hospital cardiac arrest (IHCA) accounts for a significant proportion of emergency department (ED) mortality, yet systematic data remain limited. The survival rates in IHCA ranges between 10 to 20%, as against out-of-hospital cardiac arrest (OHCA) the survival rate was 14.6% (5-7). Despite advancements in resuscitation techniques and emergency medical services (EMS), survival rates remain alarmingly low, with only a small percentage of individuals achieving full neurological recovery (7,8).

The impact of cardiac arrest extends beyond mortality statistics. Survivors often experience a range of physical, cognitive, and psychological impairments, significantly affecting their quality of life and placing a substantial burden on healthcare systems and families (9,10). This considerable public health burden underscores the critical need for continued research into prevention strategies, improved resuscitation techniques, and optimized post-cardiac arrest care to enhance survival rates and improve long-term outcomes for those affected by this devastating condition (11-15). This study aims to evaluate these predictors of mortality in a large tertiary care ED in South India, thereby

Corresponding Author: Priya Ganesan MD, Christian Medical College Vellore, Department of Emergency Medicine, Vellore, India
E-mail: priyajyotsnav@gmail.com **ORCID ID:** orcid.org/0000-0001-6542-3766

Cite this article as: John LA, Nalatha VT, Devasir GK, Ganesan P. A retrospective study of prognostic indicators of mortality in in-hospital cardiac arrest in a tertiary care emergency center. Eurasian J Emerg Med. 2026;25: 288-93.

Received: 13.12.2025
Accepted: 31.03.2026
Published: 08.04.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

contributing region-specific evidence to the global literature on cardiac arrest outcomes.

Materials and Methods

Study Design and Setting

This is a retrospective observational study conducted in the ED, a 75-bed unit of a large tertiary care referral center in Tamil Nadu, South India. The ED caters to approximately 90,000 patients annually, including both trauma and non-trauma cases. This study was approved ethics committee approval was obtained from the Institutional Review Board (IRB)/Ethics Committee operating under Christian Medical College Vellore (IRB min no: 0624160, date: 12.06.2024). Patient consent was waived, and confidentiality was maintained using unique identifiers and password-protected data entry software, which was accessible to a limited number of users.

Study Period

The study was conducted over a period of one year, from January 2023 to December 2023.

Participants

All patients aged 18 years and above who presented to the ED and sustained cardiac arrest were included in the study. Patients with OHCA, incomplete charts, or missing data were excluded from the analysis.

Variables

Data were gathered from the hospital's electronic database and recorded on a standardized data abstraction sheet. Information collected included baseline characteristics, rhythm at presentation, cause of cardiac arrest (5H/5T), history of primary treatment at another medical center, use of defibrillation, time taken to achieve return of spontaneous circulation (ROSC), presenting complaints, vital signs at presentation, and cardiac arrest outcomes in the ED and hospital.

Bias

Due to the retrospective nature of the study, neither exposure nor outcome assessment could be controlled.

Statistical Analysis

Continuous variables were expressed as means with standard deviations, and nominal variables as numbers and percentages. The chi-squared test or Fisher's exact test was used to compare variables with survival outcomes. Statistical significance was defined as a p-value less than 0.05. Logistic regression analysis was performed for significant outcome variables, with results expressed as 95% confidence intervals. Data were entered into Microsoft Excel (version 16.65), and statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS) for Windows (IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0, Armonk, NY).

Results

During the one-year study period (January 2023-December 2023), our ED received a total of 66,012 patients, including 10,422 trauma victims and 55,590 non-trauma patients. Among them, 267 (0.4%) patients were included after excluding patients with out-of-hospital cardiac arrest, missing charts, and incomplete data (Figure 1).

Baseline Demographic Characteristics and Vital Signs at Presentation

A total of 267 participants were enrolled in the study (Figure 1). The mean age of the study population was 52±17.37 years. There were 187 (70%) males and 80 (30%) females, with a male-to-female ratio of 2.3:1. The hemodynamic parameters at presentation included a mean heart rate of 97±54 bpm, mean oxygen saturation of 81±17%, mean systolic blood pressure of 117±36 mmHg, and mean respiratory rate of 32±9 breaths per minute.

Approximately 249 (93.3%) patients stayed in the ED for less than 6 hours, while 18 (6.7%) had a longer stay of more than 6 hours. The major comorbidities observed were diabetes mellitus (44%), systemic hypertension (42%), malignancy (8%), chronic kidney disease (10.8%), and ischemic heart disease (20.9%) (Table 1).

The most common initial rhythm observed during cardiovascular and pulmonary resuscitation (CPR) in our study population was pulseless electrical activity (PEA) (64%), followed by asystole (23%), ventricular fibrillation (VF) (6.7%), and ventricular tachycardia (5.6%). The defibrillator was used in 76 (28.5%) participants for shockable rhythms. The mean duration of CPR was 18±11.5 minutes after CPR, 240 participants required intubation. Hypoxia was the most common reversible cause of cardiac arrest (43%), followed by coronary thrombosis in 69 (25.8%), acidosis in 29 (10.9%), and hypovolemia in 29 (10.9%). None of the participants experienced cardiac tamponade or hypothermia.

Although 214 (80%) participants achieved ROSC, it was non sustained in 103 (48%) cases. Only 111 (52%) participants attained sustained ROSC. Out of which 29 (10.86) went to LAMA (Table 2). The majority of the study population had a glasgow coma scale score of 3-8 post-CPR. The modified Rankin score (mRS) was calculated at hospital discharge. Only 22 out of 111 had a mRS of <3, a favorable neurological outcome. About 156 (58%) patients died following cardiac arrest.

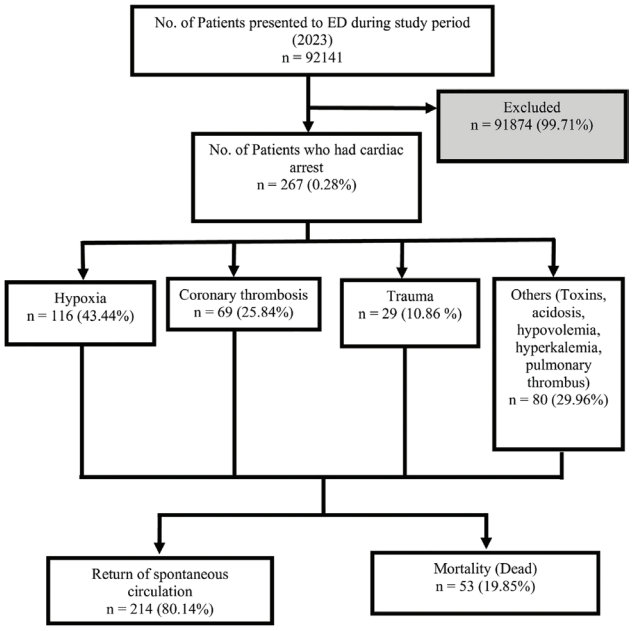


Figure 1. STROBE diagram
ED: Emergency department

Table 1. Baseline characteristics in patients		
Variable		**n=267 (%)**
Age	<40 years	95 (35.58)
	>40 years	172 (64.41)
Gender	Male	187 (70.04)
	Female	80 (29.96)
	Others	0
Duration of ED stay*	Less than 6 hours	249 (93.26)
	More than 6 hours	18 (6.74)
Heart rate	60-100	50 (18.73)
	<60	126 (47.19)
	>100	91 (34.08)
Systolic blood pressure	<90 mmHg	72 (26.96)
	>90 mmHg	195 (73.03)
Respiratory rate	>22	123 (46.06)
	<22	144 (53.93)
Saturation	<80%	165 (61.80)
	>80%	102 (38.20)
qSOFA	<2	32 (11.98)
	>2	235 (88.01)
Comorbidities	Diabetes mellitus	109 (40.82
	Systemic hypertension	105 (39.32)
	Chronic kidney disease	29 (10.86)
	Malignancy	24 (8.98)
Reversible cause¥	Thrombosis (coronary)	69 (25.84)
	Thrombosis (pulmonary)	3 (1.12)
	Trauma	29 (10.86)
	Toxins	5 (1.87)
	Tamponade (cardiac)	0
	Hydrogen ions (acidosis)	22 (8.23)
	Hypoxia	116 (43.44)
	Hypovolemia	19 (7.11)
	Hyperkalemia	19 (7.11)
	Hypothermia	0
Intubated at presentation		17 (6.36)
DNI/DNR		32 (11.99)

¥n=282: Multiple causes per patient were identified, *ED: Emergency department, qSOFA: Quick sequential organ failure assessment, DNI/DNR: Do not intubate or do not resuscitate

Table 2. ED outcome	
Outcome	**No. of patients (percentage)**
Admitted	127 (47.56)
LAMA*	87 (32.58)
Dead	53 (19.85)

*: LAMA: Leave against medical advice, ED: Emergency department

ED Outcome Analysis

Logistic regression analysis identified failure to achieve ROSC as a strong independent predictor of mortality, with an odds ratio (OR) approaching 28 [OR=27.876 (6.27-123.90), p<0.001]. Prolonged cardiopulmonary resuscitation (CPR) duration also emerged as a significant risk factor, with patients undergoing more than 20 minutes of CPR exhibiting a more than five-fold increase in adjusted odds of mortality [OR=5.109 (1.903-13.708), p=0.001]. Conversely, a primary diagnosis of acute coronary syndrome (ACS) demonstrated a protective association [OR=0.132 (0.034-0.509), p=0.021], reinforcing the established notion that cardiac aetiologies are associated with more favourable outcomes in successfully resuscitated patients. Additionally, early intubation [OR=21.89 (4.88-98.98) p<0.001] was linked to markedly improved outcomes, with a 21-fold increase in the odds of survival compared to other interventions (Table 3).

Table 3. Risk factors for cardiac arrest on multivariate analysis using logistic regression with mortality as outcome variable							
Variable alive n=214 (%)		Mortality		Unadjusted analysis		Adjusted analysis	
		Dead n=53 (%)	OR (95% CI)	p-value	OR (95% CI)	p-value	
Age	<40 yrs	28 (25.2)	42 (26.9)	1.092 (0.627,1.903)	0.756	2.017 (0.841, 4.839)	0.117
	>40 yrs	83 (74.8)	114 (73.1)				
Gender	Male	80 (72.1)	107 (68.6)	1.892 (1.102,3.251)	0.541		
	Female	31 (27.9)	49 (31.4)				
Diagnosis	ACS	27 (24.3)	59 (37.8)	0.132 (0.034,0.509)	0.021		
	Others	84 (75.7)	97 (62.2)				
ED stay	≤6 hr	109 (98.2)	140 (89.7)	6.229 (1.402,27.668)	0.016		
	≥6 hr	2 (1.8)	16 (10.3)				
Inotropes	Yes	102 (91.1)	121 (77.6)	6.836 (3.125,14.958)	.000		
	No	9 (8.1)	35 (22.4)				
Duration of CPR	≤20 min	99 (89.2)	88 (56.4)	6.375 (3.238,12.554)	0.001	5.109 (1.903, 13.708)	0.001
	≥20 min	12 (10.8)	68 (43.6)				
ROSC attained	Yes	110 (99.1)	104 (66.7)	38.653 (12.083, 123.504)	<0.001	27.876 (6.273, 123.902)	<0.001
	No	1 (0.9)	52 (33.3)				
Intubation	Yes	110 (99.1)	130 (83.3)	21.989 (4.888, 98.988)	<0.001		
	No	1 (0.9)	26 (16.7)				

ACS: Acute coronary syndrome, ED: Emergency department, CPR: Cardiovascular and pulmonary resuscitation, ROSC: Return of spontaneous circulation, OR: Odds ratio, CI: Confidence interval

Discussion

The mean age of the study population (52±17.37 years) and the prominent male predominance (70%) are consistent with demographic trends seen in global cardiac arrest registries (16,17). More critically, the high prevalence of major comorbidities, including diabetes mellitus (44%) and systemic hypertension (42%), signals a patient population with substantial pre-existing cardiovascular risk. This co-morbidity burden is a well-documented driver of poorer outcomes following critical events, complicating successful resuscitation and post-arrest recovery (18).

Hypoxia is the most common reversible cause (43%), underscoring that the primary mechanism of arrest in this setting is often an overwhelming physiological insult rather than a primary electrical event. The high mean respiratory rate (32±9 breaths per minute) and significantly low mean oxygen saturation (81%±17%) at presentation further support this interpretation, suggesting that many patients were in profound respiratory distress immediately prior to arrest.

The observed epidemiological profile of this study cohort, characterized by a staggering prevalence of PEA and asystole as initial rhythms, strongly reflects the pattern associated with medical-etiology cardiac arrests occurring within the hospital or ED setting, often secondary to respiratory failure or shock these contrasts sharply with the higher incidence of VF typically seen in community-witnessed OHCA (19).

While the achievement of initial ROSC in a high proportion of participants (80%) is technically impressive, the finding that non-sustained ROSC occurred in nearly half of these cases (48%) highlights a crucial challenge. This high rate of hemodynamic collapse in the immediate post-resuscitation phase, leading to an overall mortality rate of 58%, is comparable to international ED-based mortality figures (19,20). It emphasizes that the quality and sustainability of the post-resuscitation state, rather than transient ROSC alone, are the true determinants of survival, necessitating rapid stabilization and targeted post-cardiac arrest care.

The strong conclusions derived from the logistic regression analysis are critical for prognostication and resource allocation. The failure to achieve sustained ROSC emerged as a powerful, independent predictor of mortality, with an OR approaching 28 [OR=27.876 (6.27-123.90), p<0.001] (20). This reaffirms the

necessity of achieving a stable post-resuscitation circulation and serves as a cornerstone criterion used in most guidelines for determining when to terminate resuscitation efforts. Similarly, a prolonged CPR duration (over 20 minutes) was identified as a significant risk factor, multiplying the adjusted odds of mortality by more than fivefold [OR=5.109 (1.903-13.708), p=0.001]. This finding supports the clinical doctrine that prolonged resuscitation in the absence of reversible causes often represents efforts directed towards non-reversible underlying pathology (21,22).

Conversely, the observation that a primary diagnosis of ACS demonstrated a protective association [OR=0.132 (0.034-0.509), p=0.021] reinforces the established fact that ACS may have appeared protective due to rapid percutaneous coronary intervention availability and the cardiac etiology being more amenable to intervention. We contrasted this with literature reporting higher mortality in ACS, carrying a comparatively higher potential for survival and favorable neurological outcome (19). Furthermore, the finding that early intubation was linked to a markedly improved outcome [OR=21.989 (4.88-98.98), p<0.001] is highly relevant given the high rate of hypoxia in this ED population. This result strongly supports an aggressive approach to definitive airway management to correct the most common reversible cause in this specific clinical environment (21-23). However, this positive association must be interpreted within the context of the ED, as the timing of pre-hospital intubation remains a subject of ongoing debate.

Finally, the poor neurological outcome, with only 22 out of 111 sustained ROSC patients achieving a favorable mRS of <3, underscores the devastating consequence of global cerebral ischemia that characterizes non-shockable rhythms and prolonged low-flow states (18,24-26). This finding emphasizes the critical need for immediate and optimal implementation of high-quality post-cardiac arrest care, including targeted temperature management and meticulous hemodynamic control, to translate sustained ROSC into a meaningful neurological survival.

Study Limitations

This retrospective study faced challenges, including missing data on variables such as the mode of arrival to the ED and bystander CPR. While it quantifies the overall burden of cardiac arrest and hospital discharge morbidity, it does not fully address long-term morbidity among CPR survivors. Further studies are needed to explore these aspects and enhance the quality of resuscitation to improve the neurological outcome of cardiac arrest survivors.

Research Implications

Future research should focus on OHCA and emphasize the importance of pre-hospital EMS in improving outcomes.

Conclusion

This study demonstrates that failure to achieve sustained ROSC and prolonged CPR duration are strong, independent predictors of mortality in IHCA. Early recognition and management of reversible causes, particularly hypoxia combined with timely airway intervention and high-quality chest compressions are essential to improving survival. Although ACS appeared protective in our cohort, this finding highlights the need for further investigation into etiology-specific outcomes.

Strengthening prehospital emergency services and promoting bystander CPR training are vital to ensure early ROSC and better neurological recovery. Future research should extend to long-term outcomes and explore strategies to sustain ROSC beyond the immediate resuscitation phase.

The findings clearly identify the necessity for targeted research to advance pre-hospital resuscitation strategies, aiming to enhance survival rates and neurological outcomes in cardiac arrest patients. Strengthening emergency response systems is essential for better prognosis.

Ethics

Ethics Committee Approval: This study was approved ethics committee approval was obtained from the Institutional Review Board (IRB)/Ethics Committee operating under Christian Medical College Vellore (IRB min no: 0624160, date: 12.06.2024).

Informed Consent: This is retrospective study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: L.A.J., P.G., G.K.D., V.T.N., Design: L.A.J., P.G., Data Collection or Processing: L.A.J., G.K.D., V.T.N., Analysis or Interpretation: L.A.J., P.G., G.K.D., Literature Search: L.A.J., V.T.N., Writing: L.A.J., P.G., G.K.D.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Perkins GD, Graesner JT, Semeraro F, Olasveengen T, Soar J, Lott C, et al. European Resuscitation Council Guideline Collaborators. European Resuscitation Council Guidelines 2021: executive summary. Resuscitation. 202;161:1-60. Epub 2021 Mar 24.
2. Sudden cardiac arrest. National Heart, Lung, and Blood Institute. Available from: https://www.nhlbi.nih.gov/health/cardiac-arrest?utm_source=chatgpt.com

3. Sudden cardiac arrest facts. American Heart Association. Available from: https://www.heart.org/en/health-topics/cardiac-arrest?utm_source=chatgpt.com
4. Sudden Cardiac Arrest Foundation. Sudden cardiac arrest statistics. Available from: https://www.sca-aware.org/about-sudden-cardiac-arrest/latest-statistics?utm_source=chatgpt.com
5. Rajaram R, Rajagopalan RE, Pai M, Mahendran S. Survival after cardiopulmonary resuscitation in an urban Indian hospital. Natl Med J India. 1999;12:51-5.
6. Pandian GR, Thampi SM, Chakraborty N, Kattula D, Kundavaram PP. Profile and outcome of sudden cardiac arrests in the emergency department of a tertiary care hospital in South India. J Emerg Trauma Shock 2016;9:139-45.
7. Patel H, Mahtani AU, Mehta LS, Kalra A, Prabhakaran D, Yadav R, et al. Outcomes of out of hospital sudden cardiac arrest in India: a review and proposed reforms. Indian Heart J. 2023;75:321-6.
8. Sudden cardiac arrest and heart attack: How to perform CPR and use an AED. UC Davis Health. Available from: https://health.ucdavis.edu/blog/cultivating-health/sudden-cardiac-arrest-and-heart-attack-how-to-perform-cpr-and
9. Tran LN, Patel J, Yang J, O'Neill C, Yin D, Nguyen R, et al. The association between post-cardiac arrest cerebral oxygenation and survival with favorable neurological outcomes: a multicenter study. Resuscitation. 2020;154:85-92. Epub 2020 Jun 13.
10. Cardiac arrest: causes and symptoms. Children's Hospital of Philadelphia. Available from: https://www.chop.edu/conditions-diseases/sudden-cardiac-arrest?utm_source=chatgpt.com
11. El-Tawil S, Waly EH. Neurological consequences of cardiac arrest. Brain Behav. 2020;10:e01776.
12. Heart and Stroke Foundation. Causes and prevention of cardiac arrest. Available from: https://www.heartandstroke.ca/heart-disease/conditions/cardiac-arrest?utm_source=chatgpt.com
13. Resuscitation Council UK. Post-resuscitation care guidelines. Available from: https://www.resus.org.uk/professional-library/2025-resuscitation-guidelines/post-resuscitation-care-guidelines?utm_source=chatgpt.com
14. Nichol G, Sayre MR, Guerra F, Poole J. Defibrillation for ventricular fibrillation: a shocking update. JACC Cardiovasc Interv. 2017;3:433-9.
15. Kudenchuk PJ, Brown SP, Daya M, Rea T, Nichol G, Morrison LJ, et al. Amiodarone, lidocaine, or placebo in out-of-hospital cardiac arrest. N Engl J Med. 2016;374:1711-22. Epub 2016 Apr 4.
16. Topjian AA, Raymond TT, Atkins D, Chan M, Duff JP, Joyner BL Jr, et al. Part 4: pediatric basic and advanced life support: 2020 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Pediatrics. 2021;147:S88-159.
17. Sandroni C, Cronberg T, Sekhon M. Post-cardiac arrest brain injury: pathophysiology, treatment, and prognosis. Intensive Care Med. 2021;47:1393-411.
18. Rittenberger JC, Raina K, Holm MB, Kim YJ, Callaway CW. Association between cerebral performance category, modified rankin scale, and discharge disposition after cardiac arrest. Resuscitation. 2011;82:1036-40. Epub 2011 Apr 13.
19. DiLibero J, Misto K. outcomes of in-hospital cardiac arrest: a review of the evidence. Crit Care Nurs Clin North Am. 2021;33:343-56. Epub 2021 Jun 16.
20. Gräsner JT, Lefering R, Koster RW, Masterson S, Böttiger BW, Herlitz J, et al. EuReCa ONE Collaborators. EuReCa ONE-27 Nations, ONE Europe, ONE Registry: a prospective one month analysis of out-of-hospital cardiac arrest outcomes in 27 countries in Europe. Resuscitation. 2016;105:188-95. Epub 2016 Jun 16.
21. Huabbangyang T, Silakoon A, Papukdee P, Klaiangthong R, Thongpean C, Pralomcharoensuk W, et al. Sustained return of spontaneous circulation following out-of-hospital cardiac arrest; developing a predictive model based on multivariate analysis. Arch Acad Emerg Med. 2023;11:e33.
22. Monsieurs KG, Nolan JP, Bossaert LL, Greif R, Maconochie IK, Nikolaou NI, et al. European Resuscitation Council Guidelines for Resuscitation 2015: section 1. Executive summary. Resuscitation. 2015;95:1-80. Epub 2015 Oct 15.
23. Adrie C, Cariou A, Mourvillier B, Laurent I, Dabbane H, Hantala F, et al. Predicting survival with good neurological recovery at hospital admission after successful resuscitation of out-of-hospital cardiac arrest: the OHCA score. Eur Heart J. 2006;27:2840-5. Epub 2006 Nov 2.
24. Hollenbeck RD, McPherson JA, Mooney MR, Unger BT, Patel NC, McMullan PW Jr, et al. The effect of out-of-hospital primary percutaneous coronary intervention for ST-segment elevation myocardial infarction on survival following out-of-hospital cardiac arrest. Resuscitation. 2013;84:62-7.
25. Bernard SA. Treatment of primary respiratory arrest in the emergency department. Emerg Med J. 2008;25:733-6.
26. Lim Z, David Majewski, Stephen Ball, Judith Finn. Pre-existing comorbidities and outcome after out-of-hospital cardiac arrest: a systematic review and meta-analysis. Resuscitation. 2019;140:108-17.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Case Report
DOI: 10.4274/eajem.galenos.2026.02328
Eurasian J Emerg Med. 2026;25: 294-7

Hypersomnolence as the Presenting Feature of Thalamic Stroke

Navami Sanjeev Rao, Viswanathan Pandurangan, Bhargavi Madhaviah Vasudev, Rajkumar Mani

Sri Ramachandra Hospital, Clinic of General Medicine, Chennai, India

Abstract

Posterior circulation stroke affecting thalamo-cortical arousal networks may present with increased sleepiness, with a lack of motor symptoms. Here, we present the case of a 41-year-old man with acute-onset hypersomnolence, without focal neurological deficits. The initial computed tomography (CT) brain was normal. Metabolic and infectious workup were negative. Magnetic resonance imaging (MRI) of brain was performed and revealed bilateral paramedian thalamic infarcts with hemorrhage and a left occipital infarct. MRI angiography demonstrated reduced flow in the posterior cerebral artery segments. This case highlights the prominent clinical feature of profound hypersomnolence in thalamic stroke and emphasizes how such patients are easily misdiagnosed when early CT imaging is unrevealing. Recognition of this pattern is critical to avoid delays in the diagnosis of thalamic infarction.

Keywords: Thalamus, hypersomnolence, posterior circulation, arousal network, stroke

Introduction

A drowsy patient with no focal deficits and a normal computed tomography (CT) brain is almost always worked up for metabolic encephalopathy, toxic or infectious causes. When focal neurological signs are absent, stroke is not often considered a differential diagnosis. In the upper pons and midbrain is located the ascending reticular activating system, which functions as a critical regulator of cortical arousal and consciousness (1). Wakefulness is maintained by interconnected brainstem, hypothalamic, and basal forebrain nuclei through pathways that connect to the cortex. Sleep is promoted by inhibitory GABAergic neurons in the ventrolateral preoptic area of the hypothalamus. These neurons suppress wake-promoting circuits through reciprocal inhibition. The thalamus is not merely a sensory relay. It is the gateway through which the brainstem maintains cortical wakefulness. Bilateral injury to this relay does not cause coma. It causes a patient who has increased sleepiness and difficulty staying awake (2). We describe a patient whose only manifestation of thalamic stroke was increased sleepiness lasting for around 20 hours.

Case Report

A 41-year-old man with no known comorbidities was brought to the emergency room with complaints of difficulty staying awake. He was functioning normally the previous day, had gone to sleep at 4 pm, and was reportedly normal before sleeping. At 11:00 am the next day, attenders tried to wake him up. He opened his eyes to calls and drifted back to sleep. There was no history of fever, seizures, vomiting, headache, trauma, or limb weakness. Attenders denied any history suggesting sedative overdose. They had gone to another hospital initially, CT brain done, showed a normal study. Capillary blood glucose recorded in the emergency room was 123 mg/dL.

On examination, he was drowsy, arousable to calls, able to obey simple commands like opening and closing of eyes, finger movements, limb movements briefly, and then goes back to sleep quickly. Pupils were bilaterally equal and reactive, and there was no restriction of eye movements. He moved all four limbs spontaneously. Deep tendon reflexes were normal bilaterally, and plantar responses were mute bilaterally. As the CT brain was normal, and the patient had no focal neurological deficits, a provisional diagnosis of metabolic encephalopathy vs. septic

Corresponding Author: Navami Sanjeev Rao MD, Sri Ramachandra Hospital, Clinic of General Medicine, Chennai, India
E-mail: navamisrao1097@gmail.com **ORCID ID:** orcid.org/0009-0008-4836-8409
Cite this article as: Rao NS, Pandurangan V, Vasudev BM, Mani R. Hypersomnolence as the presenting feature of thalamic stroke. Eurasian J Emerg Med. 2026;25: 294-7.

Received: 04.03.2026
Accepted: 09.04.2026
Published: 22.04.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

encephalopathy was kept, and the patient was worked up for the same. Comprehensive laboratory evaluation was done. Hemoglobin was 12.4 g/dL, white blood cell counts were 9860 cells/mm^3, platelet count was 3.24 lakhs/mm^3. As there was no neutrophilic leukocytosis, bacterial infection was less likely. Renal profile showed serum creatinine of 0.8 mg/dl and BUN of 9 mg/dL. Serum electrolytes were also within the reference range-sodium 136 mmol/L, potassium 4 mmol/L, chloride 102 mmol/L, bicarbonate 21 mmol/L, calcium 8.8 mg/dL, magnesium 2.1 mg/dL, phosphate 2.8 mg/dL. Glycemic status assessed-HbA1c was 6.3%. Liver function test and thyroid function tests were unremarkable. Electrocardiogram demonstrated normal sinus rhythm with transthoracic echocardiogram revealing good left ventricular function and no regional wall motion abnormality. Because the clinical picture was unexplained, an magnetic resonance imaging (MRI) brain was obtained. MRI brain including diffusion weighted imaging and apparent diffusion coefficient revealed acute hemorrhagic infarct of bilateral paramedian thalami, with non-hemorrhagic infarct in the left occipital and posterior temporal lobes (Figures 1, 2). Magnetic resonance angiogram demonstrated decreased signal intensities in P3 and P4 segments of left posterior cerebral artery (PCA). Magnetic resonance venogram showed altered signal intensities in the straight sinus and left transverse sinus with corresponding T1 hyperintensities. A CT angiogram confirmed low contrast opacification of P3 and P4 segments of the left PCA. A CT venogram was done which showed no evidence of cerebral venous thrombosis. A repeat CT brain was done 2 days later, which showed a new hemorrhage in the left occipital lobe infarct. This pattern indicated posterior circulation ischemia involving thalamic perforators and cortical PCA branches rather than an isolated perforator infarct.



Figure 1. Axial DWI images showing restricted diffusion at the bilateral paramedian thalami (arrows)

DWI: Diffusion-weighted imaging

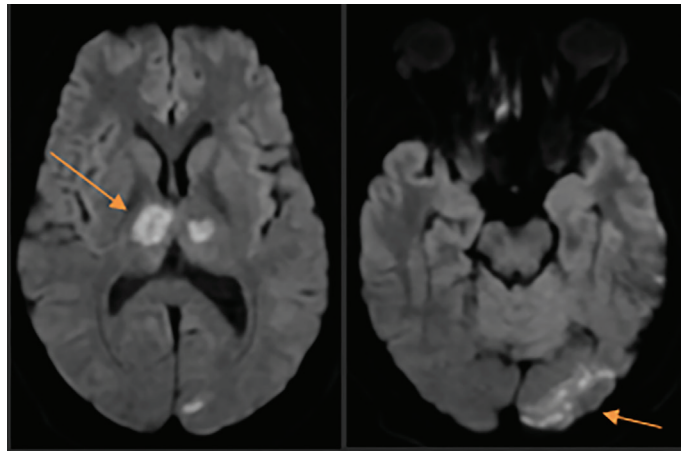


Figure 2. Axial DWI images showing restricted diffusion in bilateral thalami and the left occipital lobe reflecting the cytotoxic edema secondary to an acute infarction (arrows)

DWI: Diffusion-weighted imaging

Hospital Course

The patient was treated as an acute ischemic stroke with secondary hemorrhagic transformation, with aspirin 75 mg OD, atorvastatin 40 mg HS, levetiracetam 500 mg BD, armodafinil 100 mg OD, and antiedema measures. Blood and urine cultures were sterile. Over the next 6 days, alertness gradually improved. A detailed medication history was obtained, and the patient reported no use of sleeping pills or other centrally acting agents. No major motor deficits developed. Stroke workup, including cardiac evaluation with 24 hr holter monitoring, antinuclear antibody by immunofluorescence, antiphospholipid antibodies were unremarkable except for dyslipidemia with elevated cholesterol of 229, triglycerides 265, low-density lipoprotein 130, and high-density lipoprotein of 32. A repeat CT brain done showed the same findings and no worsening. The patient was discharged after 5 days for outpatient follow-up. The most striking feature throughout hospitalization remained excessive sleepiness out of proportion to other neurological findings.

Timeline of Clinical Events

Day 0: The patient went to sleep at 4 pm and was reportedly normal before sleeping

Day 1: Family noted difficulty waking the patient at 11 am. Visit to hospital, CT brain done, normal. Patient brought to our emergency department. After initial workup, MRI brain done showed bilateral thalamic infarct with secondary hemorrhagic transformation.

Days 1-5: Patient treated as acute ischemic stroke with secondary hemorrhagic transformation. Gradual improvement in alertness observed.

Day 5: Patient discharged with outpatient follow-up.

Follow-up and Outcomes

At outpatient follow-up, the patient reported significant improvement in daytime alertness and was able to resume routine daily activities without neurological deficits.

Discussion

Acute hypersomnia from a is not normal and should alert towards a potential stroke, making timely intervention important to patient's outcome (3). The thalamus is a deep grey matter structure on either side of the lateral wall of the third ventricle. It plays an important role in neurological functioning by acting as a transmission station for signals from nearly all pathways in the body. The thalamus participates in sensory processing, motor integration, cognition, and regulation of arousal (4). Damage to the thalamus causes impaired consciousness (42%), vertical gaze palsy (65%), memory issues (58%), and confusion (53%). If the corticospinal tract is involved, the patient may also develop weakness. Owing to its arterial connections with the brainstem, bilateral thalamic infarcts may cause cranial nerve palsies (5). The paramedian regions of the thalamus receive blood supply from small perforating branches of the first segment of the posterior cerebral arteries PCA (P1) bilaterally. In most individuals, each side is supplied independently. However, anatomical variations exist in which a single arterial trunk supplies both paramedian thalami, called the artery of percheron. The exact prevalence of this anatomical variant is not known; however, bilateral paramedian thalamic infarcts account for only 0.1% to 2% of ischemic strokes (6).

The intralaminar and dorsomedial nuclei of the thalamus serve as relays between the ascending reticular activating system in the brainstem and the cerebral cortex. Bilateral paramedian thalamic infarction disconnects the cortex from arousal input, and the patient gets a "sleeping stroke." A major diagnostic challenge arises from the radiology perspective that early CT imaging is frequently normal in paramedian thalamic infarction. Such patients closely resemble those with metabolic or toxic encephalopathy. Several reports have highlighted delays in diagnosis, leading to evaluation for other causes before a definitive MRI was obtained (7,8). While bilateral thalamic infarcts are often attributed to artery of Percheron occlusion, the occipital infarct and PCA flow abnormalities in this patient indicate a broader posterior circulation ischemic process involving both perforator and cortical branches (9,10).

Despite extensive bilateral thalamic infarction, the National Institutes of Health Stroke Scale score (NIHSS) at presentation was retrospectively calculated to be 1. This underscores a well-recognized limitation of the NIHSS in posterior circulation strokes, where deficits related to arousal, cognition, and brainstem function may be underrepresented.

Conclusion

Bilateral thalamic strokes are rare because of dual arterial blood supply. Early CT often fails to detect these infarcts. In emergency settings, it can lead clinicians towards alternative diagnoses, delaying definitive management. Persistent hypersomnolence without focal neurological deficits should prompt consideration of thalamic infarction and early MRI evaluation. This can directly influence early correct management and eligibility for reperfusion therapy.

Patient Perspective

The patient reported progressive improvement in alertness during hospitalization and expressed relief after the diagnosis was established and appropriate treatment was initiated.

Ethics

Informed Consent: Written informed consent was obtained from the patient for publication of this case report and accompanying images. The patient's identity has been protected in accordance with ethical standards.

Footnotes

Authorship Contributions

Surgical and Medical Practices: N.S.R., V.P., B.M.V., R.M., Concept: N.S.R., V.P., B.M.V., R.M., Design: N.S.R., V.P., B.M.V., R.M., Data Collection or Processing: N.S.R., V.P., B.M.V., R.M., Analysis or Interpretation: N.S.R., V.P., B.M.V., R.M., Literature Search: N.S.R., V.P., B.M.V., R.M., Writing: N.S.R., V.P., B.M.V., R.M.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Edlow JA, Rabinstein AA, Traub SJ, Wijdicks EF. Diagnosis of reversible causes of coma. Lancet. 2014;384:2064-76.
2. Schmahmann JD. Vascular syndromes of the thalamus. Stroke. 2003;34:2264-78.
3. Alattar M, Nguyen T. 0977 bilateral thalamic strokes presenting as acute hypersomnia. SLEEP. 2023;46(Suppl 1):A431.
4. Bordes S, Werner C, Mathkour M, McCormack E, Iwanaga J, Loukas M, et al. Arterial supply of the thalamus: a comprehensive review. World Neurosurg. 2020;137:310-8.
5. Sheikh Hassan M, Osman Sidow N, Adam BA, Adani AA. Superior alternating hemiplegia (Weber's syndrome): case report. Ann Med Surg (Lond). 2022;77:103674.

6. Hawkes MA, Arena JE, Rollán C, Pujol-Lereis VA, Romero C, Ameriso SF. Bilateral paramedian thalamic infarction. Neurologist. 2015;20:89-92.
7. Phate N, Pawar T, Andhale A, Singh R, Talwar D, Acharya S, et al. Artery of percheron stroke: a case report with a diagnostic challenge. Cureus. 2022;14:e21939.
8. Rajapakshe I, Senanayake B. Paramedian thalamic syndrome due to artery of percheron infarction effectively treated with early anticoagulation: a series of 3 cases. J Neurol Sci. 2019;405:62-3.
9. Naum R, Alattar M. A case of sudden onset of hypersomnolence. Annals of Internal Medicine: Clinical Cases. 2025;4.
10. Mnaili M, Bourazza A. Isolated hypersomnia due to bilateral thalamic infarcts: a case report. Int J Psychiatr Res. 2024;7:1-2.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Case Report
DOI: 10.4274/eajem.galenos.2026.74507
Eurasian J Emerg Med. 2026;25: 298-301

To Give or Not to Give: Thrombolysis for Massive Pulmonary Embolism in a Patient with Profound Anaemia

Corinne Yee Lyn Lau[1], Young Koh[1], Sameera Ganti[2], Derek Jiun Yi Heng[3], Eunizar Omar[2]

[1]Sengkang General Hospital, Clinic of Emergency Medicine, Singapore
[2]University of Malaya Medical Centre, Clinic of Emergency Medicine, Kuala Lumpur, Malaysia
[3]Woodland Health Campus, Clinic of Emergency Medicine, Singapore

Abstract

Massive pulmonary embolism (PE) is uncommon but potentially fatal, requiring immediate thrombolytic therapy. Thrombolysis, however, carries a significant bleeding risk and is generally contraindicated in cases of active bleeding, which is an important differential to consider in patients with severe anaemia. Physicians must balance possible bleeding against the risk of rapid clinical deterioration. We report a case of autoimmune haemolytic anaemia presenting with massive PE and profound anaemia successfully treated with thrombolysis.

Keywords: Emergency, critical care, respiratory, pulmonary embolism, autoimmune haemolytic anaemia, thrombolysis

Introduction

Severe dyspnea is a common presenting complaint to the emergency department (ED), with the most common causes being lower respiratory tract infection, heart failure and chronic airway disease (1). Massive pulmonary embolism (PE) and anaemia are less common causes of dyspnea, and their co-occurrence complicates diagnosis and management.

Autoimmune haemolytic anaemia (AIHA) is a rare disease where antibodies rapidly destroy red blood cells. It is usually acquired and affects 1 to 3 cases per 100,000 persons per year. AIHA can occur at any age, but its incidence increases with age (2). It has been described as an under-recognized cause of venous thromboembolism (VTE). Various mechanical and chemical processes have been proposed to explain how AIHA causes a pro-thrombotic state (3). Acute PE has been recognized as early as 1967 to be the leading cause of death in AIHA patients (4).

Emergency thrombolytic treatment in massive PE can be life-saving. However, one of the absolute contraindications for thrombolytic therapy is active bleeding, which is a differential to consider in a hypotensive patient with severe anaemia. That said, anaemia alone is neither an absolute nor relative contraindication to thrombolysis (5).

We describe a case of massive PE and profound anaemia secondary to AIHA and the considerations for thrombolytic therapy in such patients.

Case Report

A 78-year-old woman presented with one week of worsening dyspnea, pallor, and lethargy, resulting in functional decline. She denied chest pain, haemoptysis, melena, or limb swelling and had no known comorbidities or risk factors for VTE.

On arrival, her temperature was 36.5 °C, blood pressure 90/50 mmHg, heart rate 108 bpm, respiratory rate 17 breaths/min, and oxygen saturation 83% (room air). She was pale and icteric but without signs of active bleeding. Cardiopulmonary and abdominal examinations were unremarkable, and rectal examination revealed no blood.

Corresponding Author: Young Koh MD, Sengkang General Hospital, Clinic of Emergency Medicine, Singapore
E-mail: kohyoung6@gmail.com **ORCID ID:** orcid.org/0009-0000-6989-4135
Cite this article as: Lau CYL, Koh Y, Ganti S, Heng DJY, Omar E. To give or not to give: thrombolysis for massive pulmonary embolism in a patient with profound anaemia. Eurasian J Emerg Med. 2026;25: 298-301.

Received: 02.03.2026
Accepted: 05.04.2026
Published: 04.05.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Electrocardiography showed sinus tachycardia. Venous blood gas revealed pH 7.38, bicarbonate 14 mmol/L, potassium 4.1 mmol/L, and lactate 10.7 mmol/L. Haemoglobin was too low to register on the point-of-care analyzer. Point-of-care ultrasound (POCUS) demonstrated a dilated right ventricle and right atrium with a D-shaped left ventricle, suggestive of right heart strain. There was no abdominal free fluid or aortic aneurysm.

Laboratory tests confirmed severe haemolytic anaemia: haemoglobin 4.5 g/dL, haematocrit 14.8%, mean corpuscular volume 131 fL, reticulocytes 13%. Bilirubin was raised (total 64 µmol/L, direct 25 µmol/L), while liver enzymes and electrolytes were normal. Troponin-T (40 ng/L) and NT-proBNP (662 pg/mL) were elevated. A chest radiograph was normal.

Computed tomography pulmonary angiography (CTPA), performed within two hours of arrival, demonstrated extensive bilateral PE with right ventricular strain. Soon after, the patient developed worsening hypotension despite fluid resuscitation. A central venous line was inserted, and intravenous noradrenaline was commenced. Packed red blood cell transfusion was started while multidisciplinary discussion ensued between the ED, intensive care, and haematology teams.

Despite resuscitative efforts, she remained in obstructive shock. Given the haemodynamic instability, the decision was made to administer systemic thrombolysis with 100 mg recombinant tPA infused over two hours. During thrombolysis, the patient received close monitoring for haemodynamics and bleeding. Blood pressure improved within an hour, allowing vasopressor reduction. Intravenous hydrocortisone (100 mg) was administered empirically.

The direct antiglobulin test later confirmed warm AIHA (IgG and C3d positive). She was treated with high-dose corticosteroids and received several transfusions. Her haemoglobin gradually stabilized, and she was discharged after seven days on tapering oral prednisolone and anticoagulation. Follow-up CTPA at six weeks showed complete resolution of emboli and normalization of right heart size.

The patient mentioned in this report died of a myocardial infarction a few months after the index visit and had no next-of-kin available for contact. As per our institutional guidelines, which are in line with COPE guidelines for case reports on deceased and anonymized patients, consent is waived in such instances.

Discussion

This case highlights critical concepts when managing an acutely breathless and hypotensive patient in the ED. While anaemia is a known cause of breathlessness, it rarely presents with hypotension in the absence of clinically evident ongoing bleeding. Such a presentation should direct the clinician to look for other potential sources of blood loss and causes of shock. The assessment of undifferentiated shock and dyspnea is supplemented by the POCUS RUSH protocol, which is now the international standard of care (6). Since thrombolysis was indicated in the context of profound anaemia, a multi-disciplinary team approach was instituted early.

AIHA increases the risk of PE by 2.6 times as compared to those without the condition (7). Retrospective data suggest that 10-25% of AIHA patients develop PE, though many cases remain unrecognized. Risk appears heightened in the context of intravascular haemolysis and transfusion requirements. Notably, mixed-type AIHA (warm + cold autoantibodies) has been associated with deep vein thrombosis/PE in at least one reported case. Similarly, cold agglutinin disease, particularly when triggered by infections or lymphoproliferative disorders, may predispose to thrombotic complications. There are also emerging reports of AIHA complicated by PE following coronavirus disease 2019, including cases involving cold agglutinin disease (8-10). Despite various hypotheses, the exact mechanism relating AIHA and a pro-thrombotic state remains unclear. The most widely accepted hypothesis suggests that the destruction of red blood cell membrane organization by autoantibodies leads to loss of membrane phospholipid asymmetry and increased expression of anionic phospholipids, especially phosphatidylserine, which in turn results in a pro-thrombotic state (11). The management of AIHA is challenging as extensive workup is needed to diagnose the exact type, with varying treatment strategies for each. There are four known types of AIHA: warm antibody AIHA, cold antibody AIHA, paroxysmal cold hemoglobinuria and Coombs negative AIHA (12). Depending on the type, steroids, rituximab, immunosuppressive drugs (e.g., cyclophosphamide) and splenectomy may be considered. Treatment of the cause is the mainstay in cases of secondary and drug-induced AIHA (13).

Workup for AIHA takes time, and the management priorities for AIHA-PE need to be focused on rapid diagnosis, stabilization, and treatment of PE. As described by the ESC guidelines, massive PE is defined by the presence of hemodynamic instability. The ESC guidelines recommend use of rtPA in cases of massive PE, given as an infusion of 100 mg over 2 hours. The guidelines also suggest cautious volume loading and early use of norepinephrine in cases of acute right heart strain or failure (14).

This patient fulfilled the obstructive shock criteria and was deemed to have a massive PE. She was given a cautious fluid bolus of 500 mL of Lactated Ringer's and norepinephrine was started early. The dilemma remained with regard to rtPA administration

in view of the severe anaemia. Massive PE is a time-critical diagnosis, but the diagnosis of AIHA was not apparent at the ED. The patient remained undifferentiated, and ongoing bleeding was a concern. To the authors' knowledge, there are no guidelines available regarding use of rtPA for PE in the context of severe anaemia, and only case reports exist (Table 1). The only related guideline, as of 2020, is by the American Society of Haematology, which recommends use of rtPA in ischemic stroke secondary to sickle cell disease (15).

In summary, i) anyone presenting with undifferentiated shock and dyspnea should have emergency assessment with POCUS as a supplement to the usual history taking and physical examination, ii) clinicians should consider the cause of anaemia after excluding ongoing massive bleeding, iii) anaemia itself is not a contraindication to the use of systemic thrombolysis in the context of a massive PE, and iv) a multidisciplinary approach is essential for patient care and safety.

Conclusion

AIHA is an uncommon but important cause of profound anaemia and a risk factor for thromboembolic disease. Its coexistence with massive PE presents a diagnostic and therapeutic challenge. This case demonstrates that in the absence of active bleeding, thrombolysis can be safely administered despite profound anaemia when PE is immediately life-threatening.

For emergency physicians, key takeaways include the role of POCUS in rapid diagnosis, the importance of early multidisciplinary consultation, and the need to view anaemia as a relative not absolute barrier to thrombolytic therapy in massive PE.

This is one of few documented cases of successful thrombolysis for massive PE in an unrecognized AIHA patient with haemoglobin as low as 4.5 g/dL. It illustrates that timely, individualized decisions

Table 1. Reported cases of autoimmune hemolytic anaemia complicated by pulmonary embolism

Author/year	Age/sex	AIHA type	HGB at presentation	Suspected trigger	PE severity	Thrombolytic agent	Outcome
Xu et al. (2), 2020	59F	Warm AIHA	9.2 g/dL	Not mentioned	Massive PE, cardiac arrest	IV alteplase (50 mg)	Survived; recovery on 3-month follow-up
Randhawa et al. (8), 2019	66F	Warm AIHA	7.2 g/dL	Triple therapy for *H pylori*	Submassive PE	rtPA (100 mg)	Died
Imataki et al. (9), 2020	66M	Mixed-type AIHA	4.1 g/dL	Not mentioned	Fulminant/massive PE	Not mentioned	Died
Solari et al. (16), 2021	75M	Warm AIHA	8.3 g/dL	Idiopathic	Segmental (right laterobasal) and bilateral subsegmental PE	Anticoagulation with rivaroxaban	Survived
Lee et al. (17), 2016	56F	Mixed-type AIHA	2.8 g/dL	Not mentioned	Bilateral subsegmental PE	Heparin & warfarin	Survived
Shiroshita et al. (18), 2023	60F	Warm AIHA (post-steroids)	5.8 g/dL	After starting methylprednisolone	Left subsegmental PE	Apixaban	Survived
Patil et al. (10), 2022	51F	Cold agglutinin disease AIHA	5.1 g/dL	COVID-19 infection	Bilateral lower lobe segmental PE	Heparin	Survived
Present case (2024)	78F	Warm AIHA (IgG + C3d)	4.5 g/dL	Idiopathic	Massive PE with obstructive shock	rtPA (100 mg over 2 h) + steroids + transfusion	Survived; complete PE resolution

Table 1 these cases demonstrate a spectrum of severity in presentation for PE patients with AIHA, with the present case being one of two reported cases of massive PE anaemia treated successfully with full-dose thrombolysis
AIHA: Autoimmune hemolytic anemia, HGB: Hemoglobin, PE: Pulmonary embolism, COVID-19: Coronavirus disease-2019

grounded in pathophysiologic reasoning and team collaboration can be lifesaving even in the most complex scenarios.

Ethics

Informed Consent: The patient mentioned in this report died of a myocardial infarction a few months after the index visit and had no contactable next-of-kin. As per our institutional guidelines, consent is waived in such instances.

Footnotes

Authorship Contributions

Surgical and Medical Practices: D.J.Y.H., Design: S.G., Analysis or Interpretation: Y.K., Literature Search: C.L.Y.L., S.G., E.O., Writing: C.L.Y.L., Y.K., S.G., D.J.Y.H.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Laribi S, Keijzers G, van Meer O, Klim S, Motiejunaite J, Kuan WS, et al. Epidemiology of patients presenting with dyspnea to emergency departments in Europe and the Asia-Pacific region. Eur J Emerg Med. 2019;26:345-9.
2. Xu J, Wang L, Chen F. Acute PE from autoimmune hemolytic anaemia. BMC Cardiovasc Disord. 2020;20:106.
3. Thachil J. Autoimmune haemolytic anaemia--an under-recognized risk factor for venous thromboembolism. Transfus Med. 2008;18:377-8.
4. Allgood JW, Chaplin H Jr. Idiopathic acquired autoimmune hemolytic anemia. A review of forty-seven cases treated from 1955 through 1965. Am J Med. 1967;43:254-73.
5. Ucar EY. Update on thrombolytic therapy in acute pulmonary thromboembolism. Eurasian J Med. 2019;51:186-90.
6. Atkinson P, Bowra J, Milne J, Lewis D, Lambert M, Jarman B, et al. International federation for emergency medicine consensus statement: sonography in hypotension and cardiac arrest (SHoC): an international consensus on the use of point of care ultrasound for undifferentiated hypotension and during cardiac arrest. CJEM. 2017;19:459-70.
7. Ungprasert P, Tanratana P, Srivali N. AIHA and VTE. Thromb Res. 2015;136:1013-7.
8. Randhawa G, Chiu CY, Suban Na Ayutthaya T. Warm AIHA with PE. Case Rep Hematol. 2019;2019:2564682.
9. Imataki O, Iseki K, Uchida S, Uemura M, Kadowaki N. Hemolysis-associated PE. Medicine (Baltimore). 2020;99:e18984.
10. Patil NR, Herc ES, Girgis M. AIHA with PE in COVID-19. Hematol Oncol Stem Cell Ther. 2022;15:213-6.
11. Zwaal RFA, Schroit AJ. RBC membrane asymmetry. Blood. 1997;89:1121-32.
12. Liebman HA, Weitz IC. AIHA. Med Clin North Am. 2017;101:351-9.
13. Hill QA, Stamps R, Massey E, Grainger JD, Provan D, Hill A, et al. The diagnosis and management of primary autoimmune haemolytic anaemia. Br J Haematol. 2017;176:395-411. Epub 2016 Dec 22.
14. Konstantinides SV, Meyer G, Becattini C, Bueno H, Geersing GJ, Harjola VP, et al. 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS). Eur Heart J. 2020;41:543-603.
15. DeBaun MR, Jordan LC, King AA, Schatz J, Vichinsky E, Fox CK, et al. American Society of Hematology 2020 guidelines for sickle cell disease: prevention, diagnosis, and treatment of cerebrovascular disease in children and adults. Blood Adv. 2020;4:1554-88.
16. Solari D, Alberio L, Ribi C, Grandoni F, Stalder G. Autoimmune hemolytic anemia and pulmonary embolism: an association to consider. TH Open. 2021;5:e8-13.
17. Lee M, Higuchi T, Koyamada R, Okada S. PE in mixed AIHA. J Hematol. 2016;5:70-3
18. Shiroshita K, Okayama M, Soma H, Sato Y, Hayashi H, Shiozawa Y, et al. Thromboembolism after steroids in AIHA. Clin Hematol Int. 2023;5:165-9.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.56514
Eurasian J Emerg Med. 2026;25: 302-7

Clinical and Demographic Characteristics, Frequency, and Reasons for Emergency Department Visits Among Patients Receiving Home Health Care

Sueda Zaman[1], Emrah Arı[1], Nureddin Servi[1], Merve Unutmaz[2]

[1]Mamak State Hospital, Clinic of Emergency Department, Ankara, Türkiye
[2]Ankara Bilkent City, Clinic of Emergency Department, Ankara, Türkiye

Abstract

Aim: This study aimed to evaluate the demographic and clinical characteristics and reasons for emergency department (ED) visits among patients receiving home health care services and to investigate the impact of comorbidity burden on 30-day ED revisits.

Materials and Methods: This retrospective study was conducted in the ED of Mamak State Hospital. Patients receiving home health care services who presented to the ED between August 2023 and December 2024 were included. Demographic and clinical data of 299 patients were obtained from hospital records. Comorbidity burden was assessed using the Charlson Comorbidity index (CCI). The primary outcome was an ED revisit within 30 days. Statistical analyses included Mann-Whitney U test, chi-square test, ROC analysis, and logistic regression.

Results: Among 299 patients, 62 (20.7%) had a 30-day ED revisit. Patients with revisits had significantly higher CCI scores than those without revisits [median 6 (4-8) vs. 5 (3.5-6); p <0.001)]. ROC analysis showed that the CCI had moderate discriminative ability to predict revisits (area under the curve= 0.626). At a cut-off value of CCI >6, specificity was 78.1% and sensitivity was 40.3%. In multivariate analysis, CCI was an independent predictor of 30-day revisits (odds ratio: 1.231; 95% confidence interval: 1.075-1.410; p=0.003). Infectious causes were more frequent among patients with revisits, although the difference was of borderline statistical significance (p=0.057).

Conclusion: Comorbidity burden is an important predictor of 30-day ED revisits among patients receiving home health care services. The CCI may help identify high-risk patients and guide closer follow-up with the aim of reducing recurrent ED visits.

Keywords: Home care services, hospital, Charlson Comorbidity index, emergency department visits

Introduction

Home health care services are an important form of healthcare delivery, enabling individuals with chronic diseases and those requiring ongoing care to receive medical treatment in their homes rather than in hospitals. These services play a vital role particularly in the management of chronic diseases and in meeting long-term care needs, especially in parallel with the increasing elderly population (1). As the aging population continues to grow, the prevalence of chronic diseases also increases, leading to a greater demand for home health care services. Therefore, it is essential to further improve the infrastructure and expand the scope of healthcare services provided within home health care systems (2).

According to the regulations, the criteria for patients eligible to receive home health care services are defined as follows:

1. Individuals who are dependent on medical devices, bed, or home due to diagnosed diseases and/or who experience difficulty accessing healthcare services because of advanced age, and who request to receive healthcare services at their place of residence and whose request is deemed appropriate.

2. The following individuals may be included within the scope of home health care services,

a) All individuals aged 80 years and older who request the service,

Corresponding Author: Sueda Zaman MD, Mamak State Hospital, Clinic of Emergency Department, Ankara, Türkiye
E-mail: drsuedazaman@gmail.com **ORCID ID:** orcid.org/0000-0002-4830-9432
Cite this article as: Zaman S, Arı E, Servi N, Unutmaz M. Clinical and demographic characteristics, frequency, and reasons for emergency department visits among patients receiving home health care. Eurasian J Emerg Med. 2026;25: 302-7.

Received: 11.03.2026
Accepted: 22.04.2026
Published: 11.05.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

b) Patients aged 65 years and older with a documented chronic disease confirmed by a medical board report and classified as fully dependent or severely dependent in terms of activities of daily living,

c) Patients who are dependent on medical devices and/or home due to their illness,

d) Patients who have received palliative care and whose medical care is considered appropriate to continue at home,

e) Patients for whom home health care services are deemed appropriate by the physician upon hospital discharge and who require continuous medical care according to a treatment plan,

f) Individuals who require medical care for up to thirty days following discharge from a healthcare facility (which may be extended for additional thirty-day periods when necessary by a physician) and who are determined to require temporary medical care at home are eligible for such care.

3. Patients with diabetes, heart failure, stroke, chronic obstructive pulmonary disease, Alzheimer's disease/dementia, neurodegenerative diseases, terminal-stage cancer, chronic diseases with multimorbidity, patients undergoing treatment and follow-up after hip/knee/shoulder fractures, and those requiring medical care due to amputation are prioritized for evaluation within the scope of home health care services (3).

The expansion of home health care services has contributed significantly to reducing hospital length of stay and lowering patient care costs. However, patients receiving these services may still present to hospitals again due to complications or inadequate access to healthcare services (4). The majority of patients receiving home health care services consist of geriatric individuals aged 65 years and older with chronic diseases (5). In this geriatric population, multiple comorbid conditions and polypharmacy are among the main factors that increase emergency department visit rates (6).

To reduce emergency department utilization, home care protocols should be tailored to patient profiles, and educational programs aimed at improving the prognosis of common chronic diseases should be implemented. Such approaches may contribute to reducing emergency department visits (7).

Although public health strategies have been implemented to increase access to home-based or community-based healthcare services in order to prevent overcrowding in emergency departments, scientific evidence regarding the relationship between home health care services and emergency department utilization remains surprisingly limited and controversial (8). In geriatric patient groups with frequent emergency department visits, the Charlson Comorbidity index (CCI) is a validated, simple, and easily applicable method used to estimate the mortality risk associated with comorbid diseases and is widely utilized as a predictor of long-term prognosis and survival (9).

In this context, the present study retrospectively examined the demographic and clinical characteristics and the reasons for emergency department visits among patients receiving home health care services. In particular, the study aimed to evaluate the impact of comorbidity burden on emergency department revisits within 30 days.

Materials and Methods

This retrospective study was conducted in the Emergency Department of Mamak State Hospital. Patients who were receiving home health care services and who presented to the emergency department between August 2023 and December 2024 were included in the study. Prior to data collection, ethical approval was obtained from the Scientific and Ethical Evaluation Committee for Medical Research No. 2 of Ankara Bilkent (approval no: TABED 2-25-1114, date: 30/04/2025). The study was conducted in accordance with the principles of the Declaration of Helsinki.

Patients who presented to the emergency department during the study period and who were registered in the home health care service program were included. Patients with missing clinical data were excluded. A total of 299 patients were included in the study.

Demographic and clinical characteristics of the patients were obtained retrospectively from hospital records, including age, sex, bedridden status, presence of pressure ulcers, and chronic comorbidities (diabetes mellitus, coronary artery disease, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, lung disease, liver disease, connective tissue disease, and malignancy). Additionally, information regarding the presence of nasogastric tube/percutaneous endoscopic gastrostomy, urinary catheter use, incontinence pad use, and oxygen concentrator use was recorded.

Furthermore, the diagnoses at the initial emergency department visit were categorized as infectious, cardiovascular, musculoskeletal, gastrointestinal, or other.

The comorbidity burden of the patients was assessed using the CCI. The primary outcome of the study was defined as a revisit to the emergency department within 30 days following the initial emergency department visit.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, Version 25.0 (IBM Corp., Armonk, NY, USA). Continuous variables were expressed as median and interquartile range, whereas categorical variables were presented as numbers and percentages.

The Mann-Whitney U test was used to compare continuous variables between patients with and without revisits, and the chi-square test to compare categorical variables.

The discriminative ability of the CCI to predict 30-day emergency department revisits was evaluated by ROC curve analysis, and the area under the curve (AUC) was calculated. The optimal cut-off value was determined according to the ROC analysis.

Univariate logistic regression analysis was performed to identify factors associated with emergency department revisits within 30 days. Variables with $p<0.10$ in the univariate analysis were included in the multivariate logistic regression model. The results were presented as odds ratio (OR) with 95% confidence interval (CI). A p value of <0.05 was considered statistically significant in all analyses.

Results

Among the 299 patients included in the study, 62 (20.7%) revisited the emergency department within 30 days. Patients who revisited the emergency department had significantly higher CCI scores compared with those who did not [median 6 (4-8) vs. 5 (3.5-6), $p<0.001$)]. However, no significant differences were observed between the groups in terms of age, sex, or most clinical variables (Table 1).

When the distribution of diagnostic categories at the initial emergency department visit was examined according to revisit status, infectious causes were more frequent in the revisit group. However, the overall comparison remained at the borderline of statistical significance ($p=0.057$) (Table 2).

ROC analysis demonstrated that the CCI had moderate discriminative ability for predicting 30-day emergency department revisits (AUC=0.626; 95% CI: 0.548-0.704; $p<0.001$) (Figure 1).

At a CCI cut-off of >6, specificity was relatively high (78.1%), whereas sensitivity was lower (40.3%) (Table 3).

In the multivariate logistic regression analysis, CCI was the only independent predictor of 30-day emergency department revisits.

Table 1. Baseline demographic and clinical characteristics of patients according to 30-day emergency department revisit

Variable	All patients n=299 (%)	No revisit within 30 days n=237 (79.3%)	Revisit within 30 days n=62 (20.7%)	p value*
Age	78 (67-84)	78 (65-84)	79 (74-84)	0.065
Female sex	200 (66.9%)	165 (69.6%)	35 (56.5%)	0.050
Bedridden status	36 (12%)	26 (11%)	10 (16.1%)	0.267
Pressure ulcer	14 (4.7%)	11 (4.6%)	3 (4.8%)	0.948
Diabetes mellitus	101 (33.7%)	73 (30.8%)	28 (45.2%)	0.034
Coronary artery disease	32 (9%)	20 (8.4%)	12 (19.4%)	0.013
Congestive heart failure	73 (24.4%)	56 (23.6%)	17 (27.4%)	0.537
Peripheral vascular disease	11 (3.7%)	7 (3%)	4 (6.5%)	0.193
Cerebrovascular disease	33 (11%)	22 (9.3%)	11 (17.7%)	0.059
Dementia	25 (8.4%)	21 (8.9%)	4 (6.5%)	0.542
Lung disease	68 (22.7%)	49 (20.7%)	19 (30.6%)	0.096
Liver disease	34 (11.4%)	24 (10.1%)	10 (16.1%)	0.186
Connective tissue disease	16 (5.3%)	9 (3.8%)	7 (11.3%)	0.020
Malignancy	20 (6.7%)	13 (5.5%)	7 (11.3%)	0.104
Nasogastric tube/percutaneous endoscopic gastrostomy	7 (2.3%)	6 (2.5%)	1 (1.6%)	0.671
Urinary catheter	14 (4.7%)	10 (4.2%)	4 (6.5%)	0.460
Incontinence pad use	115 (38.5%)	92 (38.8%)	23 (37.1%)	0.804
Charlson Comorbidity index	5 (4-7)	5 (3.5-6)	6 (4-8)	<0.001
Oxygen concentrator use	12 (4%)	11 (4.6%)	1 (1.6%)	0.280

*Continuous variables are presented as medians (IQRs) The Mann-Whitney U test was used for continuous variables, and the χ^2 test was used for categorical variables

Table 2. Association between diagnostic categories at the ınitial emergency department visit and 30-day revisit			
Diagnosis category	**No revisit within 30 days n=237 (79.3%)**	**Revisit within 30 days n=62 (20.7%)**	**p value***
Infectious	65 (27.4%)	23 (37.1%)	0.057
Cardiovascular	45 (19%)	17 (27.4%)	
Musculoskeletal	66 (27.8%)	11 (17.7%)	
Gastrointestinal	26 (11%)	8 (12.9%)	
Other (allergy, migraine, vertigo, etc.)	35 (14.8%)	3 (4.8%)	
*Chi-square test			

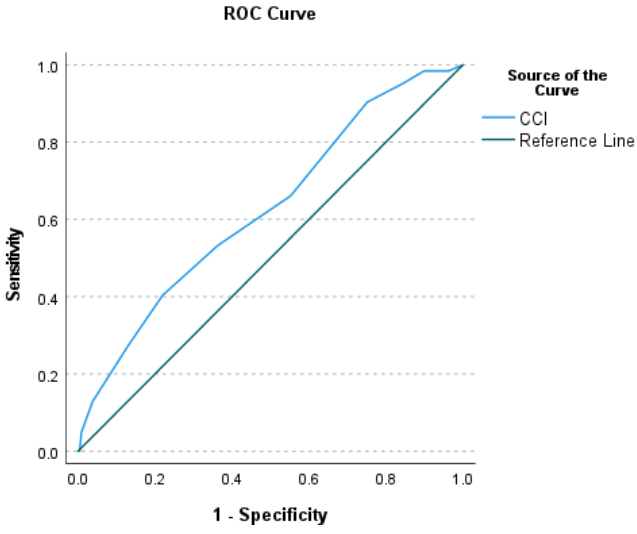


Figure 1. ROC curve for the Charlson Comorbidity index in predicting 30-day emergency department revisit (AUC=0.626)

AUC: Area under the curve, CCI: Charlson Comorbidity index

Each one-point increase in CCI was associated with approximately a 23% increase in the likelihood of a revisit (OR: 1.231; 95% CI: 1.075-1.410; p=0.003) (Table 4).

Discussion

In this study, the rate of emergency department revisits within 30 days among patients receiving home health care services was found to be 20.7%. Studies evaluating emergency department revisit rates in geriatric populations have reported that the revisit rate generally ranges between 15% and 25%. In particular, short-term emergency department revisits are more frequently observed among elderly patients with a high burden of chronic diseases (10-12). Similarly, studies examining emergency department utilization among geriatric patients have shown that older adults have a high rate of recurrent emergency department visits, which constitutes a significant burden on the healthcare system (13,14). Recent studies have also demonstrated that emergency department revisits in the geriatric population not only increase healthcare system costs but are also associated with a higher risk of hospitalization and mortality (15,16).

In our study, CCI scores were found to be significantly higher among patients who revisited the emergency department. The CCI is a prognostic tool developed to measure comorbidity burden and is widely used in clinical research (17). Although it was initially developed to predict mortality, subsequent studies have shown that it is also associated with various clinical outcomes, including hospitalization, healthcare utilization, and emergency department revisits (18). Particularly in elderly patient groups, an increased comorbidity burden leads to a reduction in physiological reserves and results in greater clinical vulnerability during acute illnesses (19). Therefore, patients with multiple chronic conditions are more likely to experience emergency department visits and recurrent admissions (20). Consistent with the existing literature, our findings support these observations. Similarly, studies conducted among patients receiving community-based healthcare services have demonstrated that comorbidity burden is one of the key determinants of emergency department utilization (8,21). The identification of CCI as an independent risk factor in our multivariate analysis further supports these findings.

In the ROC analysis, the CCI demonstrated a moderate discriminative ability in predicting emergency department revisits (AUC=0.626). Several clinical risk models used in geriatric emergency department populations have also been reported to have a moderate predictive performance (16). One possible

Table 3. Diagnostic performance of the Charlson Comorbidity index in predicting 30-day emergency department revisits (ROC analysis)										
Variable	**AUC**	**95% CI**	**Cut-off value**	**Sensitivity (%)**	**Specificity (%)**	**PPV**	**NPV**	**LR+**	**LR-**	**p value**
CCI	0.626	0.548-0.704	>6	40.32	78.06	32.5	83.3	1.84	0.76	<0.001
CCI: Charlson Comorbidity index, PPV: Positive predictive value, NPV: Negative predictive value, LR+: Positive likelihood ratio, LR-: Negative likelihood ratio, AUC: Area under the curve, CI: Confidence interval										

Table 4. Univariate and multivariate logistic regression analyses for predictors of 30-day emergency department revisit				
Variable	**Univariate OR (95% CI)**	**p value**	**Multivariate OR (95% CI)**	**p value**
Charlson Comorbidity index	1.250 (1.095-1.426)	<0.001	1.231 (1.075-1.410)	0.003
Sex	1.768 (0.996-3.136)	0.051	1.678 (0.921-3.057)	0.091
Diagnosis at initial visit	1.068 (0.513-2.222)	0.075	1.034 (0.484-2.209)	0.218
Bedridden status	1.561 (0.708-3.438)	0.269		
Pressure ulcer	1.045 (0.282-3.865)	0.948		
Nasogastric tube/PEG	0.631 (0.075-5.341)	0.673		
Urinary catheter	1.566 (0.474-5.171)	0.462		
Incontinence pad use	0.929 (0.522-1.656)	0.804		
CI: Confidence interval, OR: Odds ratio, PEG: Percutaneous endoscopic gastrostomy				

explanation is that clinical outcomes in the geriatric population are influenced not only by comorbidity burden but also by multiple factors such as functional status, cognitive status, social support, and care needs (19,20). Therefore, future risk models incorporating functional status and social determinants of health may achieve higher predictive accuracy.

Diabetes and cardiovascular diseases are among the most common chronic conditions in the geriatric population and significantly increase healthcare utilization (9). Similarly, in our study diabetes mellitus and coronary artery disease were more prevalent among patients receiving home health care services who presented to the emergency department. When diagnoses at the initial visit were evaluated, infectious causes were more frequently observed among patients who revisited; however, the difference remained of borderline statistical significance. Nevertheless, infections are well known to be one of the major causes of emergency department visits among geriatric patients. Age-related decline in immune system function and the presence of chronic diseases increase susceptibility to infections (22). In addition, infections in geriatric patients with a high comorbidity burden may lead to functional deterioration and increased care needs, thereby increasing healthcare utilization (13).

In our study, no significant association was found between either age or sex and emergency department revisits. Although some studies have reported that advanced age and male sex may increase healthcare utilization, these findings have not been consistently observed across all studies (12,16). Existing evidence suggests that in geriatric populations, clinical outcomes are more strongly influenced by factors such as comorbidity burden, functional dependency, and care needs rather than demographic characteristics (19,20).

Our findings suggest that systematic evaluation of comorbidity burden among patients receiving home health care services is clinically important. Identifying patients with high CCI scores as high-risk during emergency department visits and planning closer follow-up for these individuals may help reduce recurrent visits. Furthermore, stronger coordination between home health care services and emergency departments may help reduce unnecessary hospital visits in this patient population (8,21).

Study Limitations

This study has several limitations. First, the retrospective design and the single-center setting may limit the generalizability of the findings. In addition, some geriatric variables such as functional status, social support, and medication burden could not be evaluated. Nevertheless, this study contributes to the literature as one of the few studies examining emergency department revisits and their relationship with comorbidity burden among patients receiving home health care services.

Conclusion

This study demonstrated that comorbidity burden is an important determinant of emergency department revisits within 30 days among patients receiving home health care services. The CCI may serve as a practical tool for identifying high-risk patients during emergency department visits. Early identification of high-risk patients using simple clinical tools such as the CCI may facilitate targeted interventions and improve healthcare resource utilization in this vulnerable population. Future multicenter prospective studies may contribute to developing effective strategies to reduce recurrent emergency department visits among patients receiving home health care services.

Ethics

Ethics Committee Approval: Prior to data collection, ethical approval was obtained from the Scientific and Ethical Evaluation Committee for Medical Research No. 2 of Ankara Bilkent (approval no: TABED 2-25-1114, date: 30/04/2025). The study was conducted in accordance with the principles of the Declaration of Helsinki.

Informed Consent: This is retrospective study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: S.Z., N.S., Concept: S.Z., N.S., M.U., Design: S.Z., E.A., Data Collection or Processing: S.Z., N.S., Analysis or Interpretation: E.A., M.U., Literature Search: S.Z., N.S., Writing: S.Z., E.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Kurt ME, Çemberlitaş İ, Şen MA. EVDE Sağlık hizmetleri birimi tarafından sunulan hizmetlerin sosyal devlet olma ilkesi açisindan incelenmesi. Dicle Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi. 2019;9:306-18.
2. Maç ÇE. Erfelek İlçe Devlet Hastanesi evde sağlıkbirimi hizmetlerinin değerlendirilmesi. Aile Hekimliği. 2018;10:1-6.
3. Republic of Türkiye Official Gazette. Regulation on the Provision of Home Health Services (Evde Sağlık Hizmeti Sunumu Hakkında Yönetmelik). Official Gazette. Available from: https://www.resmigazete.gov.tr/eskiler/2023/06/20230602-1.htm.
4. Zeynep Karaali, Beste Tacal Aslan, Mehmet Burak Aktuğlu, Macit Koldaş, Elif Sinem İplik, Bedia Çakmakoğlu. Evde sağlık hizmeti veren hastaların laboratuvar özellikleri. Denendi. 2019;9:16-22.
5. İncesu E, Tombul C, Arkan M, Babuçcu H. Evde sağlık hizmetlerin erişilebilirlik: Konya Seydişehir Devlet Hastanesi evde sağlik hizmetleri birimi örneği. Tıbbi Sosyal Hizmet Dergisi. 2014;4:18-25.
6. Öztürk GZ, Toprak D. Sociodemographic characteristics of patients registered with a home care unit and an evaluation of the health services offered. Sisli Etfal Hastan Tip Bul. 2018;52:41-6.
7. Kıskaç, N. and B.J.Y.B.H.D. Girgin, Evde Sağlık hizmetlerinin dünü, Bugünü ve hemşirelerin rolü. 202;26:119-23.
8. Chiu WY, Yeh TC, Yang CC. Factors associated with emergency department visits among patients receiving publicly-funded homecare services: a retrospective chart review from Southern Taiwan Regional Hospital. Int J Health Policy Manag. 2023;12:7377.
9. Radovanovic D, Seifert B, Urban P, Eberli FR, Rickli H, Bertel O, et al. Validity of charlson comorbidity index in patients hospitalised with acute coronary syndrome. Insights from the nationwide AMIS Plus registry 2002-2012. Heart. 2014;100:288-94.
10. McCusker J, Cardin S, Bellavance F, Belzile E. Return to the emergency department among elders: patterns and predictors. Acad Emerg Med. 2000;7:249-59.
11. Friedmann PD, Jin L, Karrison TG, Hayley DC, Mulliken R, Walter J, et al. Early revisit, hospitalization, or death among older persons discharged from the ED. Am J Emerg Med. 2001;19:125-9.
12. Aminzadeh F, Dalziel WB. Older adults in the emergency department: a systematic review of patterns of use, adverse outcomes, and effectiveness of interventions. Ann Emerg Med. 2002;39:238-47.
13. Samaras N, Chevalley T, Samaras D, Gold G. Older patients in the emergency department: a review. Ann Emerg Med. 2010;56:261-9.
14. Gruneir A, Silver MJ, Rochon PA. Emergency department use by older adults: a literature review on trends, appropriateness, and consequences of unmet health care needs. Med Care Res Rev. 2011;68:131-55.
15. Rising KL, Victor T, Hollander J, Carr BG. Patient returns to the emergency department: the time-to-return curve. Acad Emerg Med. 2014;21:864-71.
16. Lowthian JA, McGinnes RA, Brand CA, Barker AL, Cameron PA. Discharging older patients from the emergency department effectively: a systematic review and meta-analysis. Age Ageing. 2015;44:761-70.
17. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40:373-83.
18. Frenkel WJ, Jongerius EJ, Mandjes-van Uitert MJ, van Munster BC, de Rooij SE. Validation of the Charlson Comorbidity index in acutely hospitalized elderly adults: a prospective cohort study. J Am Geriatr Soc. 2014;62:342-6.
19. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001;56:M146-56.
20. Salvi F, Morichi V, Grilli A, Giorgi R, De Tommaso G, Dessì-Fulgheri P. The elderly in the emergency department: a critical review of problems and solutions. Intern Emerg Med. 2007;2:292-301.
21. Campagna S, Borraccino A, Politano G, Benso A, Dalmasso M, Dimonte V, et al. Determinants associated with the risk of emergency department visits among patients receiving integrated home care services: a 6-year retrospective observational study in a large Italian region. Int J Health Policy Manag. 2021;10:605-12.
22. Yoshikawa TT. Epidemiology and unique aspects of aging and infectious diseases. Clin Infect Dis. 2000;30:931-3.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.12058
Eurasian J Emerg Med. 2026;25: 308-14

Predicting Mortality in Gastrointestinal Bleeding: Serum Copeptin Levels or Age, Systolic Blood Pressure and Shock Index-A Cross-Sectional Study

Emrah Akın[1], Alten Oskay[2], Özgen Kılıç Erkek[3], Merve Akın[4], Mert Özen[2], Atakan Yılmaz[2], Murat Seyit[2], Orhan Tamer Ereren[5], Melek Bor Küçükatay[3], İbrahim Türkçüer[2]

[1]Servergazi State Hospital, Clinic of Emergency Medicine, Denizli, Türkiye
[2]Pamukkale University Faculty of Medicine, Department of Emergency Medicine, Denizli, Türkiye
[3]Pamukkale University Faculty of Medicine, Department of Physiology, Denizli, Türkiye
[4]Republic of Turkey, Van Provincial Health Directorate, Clinic of Public Health, Van, Türkiye
[5]Bartın State Hospital, Clinic of Emergency Medicine, Bartın, Türkiye

Abstract

Aim: Vital signs, blood parameters, and some blood biomarkers in patients with gastrointestinal (GI) bleeding vary depending on the severity of the bleeding. The aim of this study is to investigate the association of vital signs and copeptin levels with transfusion requirements, and patient prognosis, including hospitalization duration and 30-day mortality.

Materials and Methods: Prospective, cross-sectional, observational study, conducted in the emergency department (ED) of a tertiary care university hospital, between June 2021 and May 2022. Admission serum copeptin levels and vital parameters of patients with a diagnosis of GI bleeding were noted and evaluated in terms of outcome measures.

Results: The study included 118 patients, 75 (63.6%) of whom were male. The median age was 72 years (interquartile range: 58-83). A total of 18 (15%) patients died within 30 days. Serum copeptin levels did not differ between deceased patients and survivors. Logistic regression analysis showed that a shock index above 0.70 and patients aged 75 years and older together [p=0.017, odds ratio (OR)=6.824; 95% confidence interval (CI): (1.419-32.811)] lead to an increase in 30-day mortality. Lower systolic blood pressure at the admission to the ED was associated with increased 30-day mortality [p=0.003, OR=0.952; 95% CI: (0.922-0.983)].

Conclusion: Although serum copeptin levels failed to predict 30-day mortality in patients admitted to the ED due to GI bleeding, our models demonstrated that age over 75 years, lower systolic blood pressure, and the shock index, were superior predictors of mortality.

Keywords: Copeptin, elderly, gastrointestinal bleeding, mortality, prognosis, shock index

Introduction

Gastrointestinal (GI) system bleeding can range from microscopic levels detectable only through laboratory tests to visible amounts detected macroscopically in feces or vomit (1,2). It frequently leads to emergency department (ED) admissions for medical interventions such as blood transfusions and endoscopic procedures. Given the significant morbidity and mortality associated with GI bleeding (1-4), understanding the factors influencing prognosis and accurately predicting outcomes in GI bleeding cases can facilitate early treatment planning and management strategies.

The blood losses due to GI bleeding cause various changes in the vital signs, blood parameters and hormonal status of patients (5). Copeptin, which is a glycosylated peptide consisting of a leucine-rich core composed of 39 amino acids, is released in equimolar amounts with the antidiuretic hormone (6,7). Studies have shown

Corresponding Author: Alten Oskay MD, Pamukkale University Faculty of Medicine, Department of Emergency Medicine, Denizli, Türkiye
E-mail: oskayten@gmail.com **ORCID ID:** orcid.org/0000-0003-4373-6280

Cite this article as: Akın E, Oskay A, Kılıç Erkek Ö, Akın M, Özen M, Yılmaz A, et al. Predicting mortality in gastrointestinal bleeding: serum copeptin levels or age, systolic blood pressure and shock index-a cross-sectional study. Eurasian J Emerg Med. 2026;25: 308-14.

Received: 21.01.2026
Accepted: 22.04.2026
Published: 13.05.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

that the release of copeptin is stimulated by processes such as hypovolemia or stress, correlates with plasma osmolarity, can be influenced by sex, remains constant with age, and does not exhibit circadian variations (7-10). The in vitro stability of copeptin and its long half-life in circulation contribute to its stable serum levels. This makes copeptin an ideal biomarker (8,9).

The relationship between copeptin and cardiovascular, pulmonary, neurological, and infectious conditions is investigated in many studies such as heart failure (11), systemic inflammatory response syndrome, sepsis, septic shock (12), and acute stroke (13). In 2013, Salt et al. (5) investigated copeptin levels in patients with acute GI bleeding and a control group and found a significant difference between the groups. However, the study did not evaluate the relationship between copeptin and prognosis.

Our study aims to assess serum copeptin levels in patients with GI bleeding and to investigate the correlation between copeptin levels, transfusion requirements, and patient prognosis, including hospitalization duration and 30-day mortality.

Materials and Methods

Ethical Statement

The ethical suitability of this research was approved at the meeting of Pamukkale University the Non-Interventional Clinical Research Ethics Committee (decision number: 10, date: 25.05.2021). Written informed consent was obtained from all participants in accordance with the ethical principles for research involving human subjects as stated in the Declaration of Helsinki.

Study Group

This prospective, cross-sectional, observational study was conducted in the ED of Pamukkale University Hospital, a tertiary care facility that receives approximately 200,000 patient visits annually. Over the course of 12 months (June 2021-May 2022), 206 patients aged 18 years and older with a preliminary diagnosis of GI bleeding were admitted to the ED and voluntarily enrolled in the study. Patients were initially evaluated according to the predefined inclusion and exclusion criteria (Figure 1). A total of 10 patients were excluded at this stage because they met the exclusion criteria. Additionally, 78 patients who were initially suspected of having GI bleeding but were subsequently ruled out by GI endoscopy were excluded (Table 1). After these exclusions, 118 patients with a final diagnosis of GI bleeding were included in the final analysis.

The inclusion criteria for the study were the age of 18 years or older and the presence of one of the following complaints: hematemesis (bloody vomiting), bloody stools, or melena. Patients or their legal guardians signed an informed consent form to participate in the study. Exclusion criteria included patients with end-stage renal failure (glomerular filtration rate <15 mL/min), patients with known diseases affecting the hypothalamic-pituitary axis (e.g., diabetes insipidus), pregnant women, patients who had received vasoactive drugs or blood or

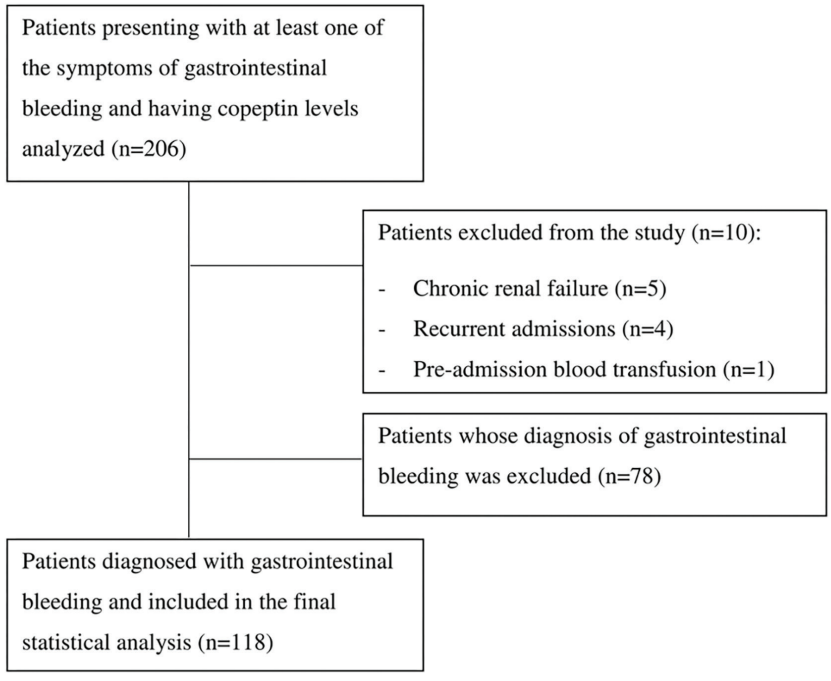


Figure 1. Flowchart of the study

blood product transfusions before admission, and patients with recurrent admissions.

Patient demographics, comorbidities, medications, symptoms at admission to the ED, symptom duration, vital signs, shock index (heart rate/systolic blood pressure), mean arterial pressure, serum copeptin levels, biochemistry results, blood gas values, GI endoscopy findings, transfusion treatments, and admission outcomes (discharge, hospitalization, mortality) were recorded on the study form. The study endpoints included duration of hospitalization and 30-day mortality.

Copeptin Measurement and Laboratory Analysis

To measure copeptin levels, 5 mL of venous blood was collected from 206 patients who presented with symptoms of GI bleeding at their first admission to the ED. Following a 10-15-min rest at room temperature, the blood samples were centrifuged at 3000 rpm for 20 min. The separated serum samples were then stored at -80 °C freezer for subsequent analysis. A double-antibody sandwich ELISA was used to measure copeptin levels in the serum. The analyses were conducted in the laboratories of the Physiology Department. Copeptin levels were measured using a human copeptin ELISA kit (Bioassay Technology Laboratory, Human Copeptin ELISA Kit, Cat. No. E1129Hu) in accordance with the manufacturer's recommendations.

Statistical Analysis

The SPSS v.25 software package (IBM SPSS Statistics 25, Armonk, NY: IBM Corp.) was used to analyze the data. Descriptive statistics included number and percentage for categorical variables, and arithmetic mean, standard deviation, median, and interquartile range (IQR) (1^{st}-3^{rd} quartile) for continuous variables. Categorical variables were compared using the chi-square test. The Kolmogorov-Smirnov and Shapiro-Wilk tests were used to assess the normality of the dataset. Since the assumptions for parametric tests were not met, the Mann-Whitney U test and Kruskal-Wallis test were used to compare independent group differences. The Spearman correlation test was used to assess the correlation between numerical variables. Backward logistic regression analysis was conducted to explore the independent effects of factors on 30-day mortality. Statistical significance was set at $p<0.05$.

Results

Of the 128 patients who received the final diagnosis of GI bleeding, 10 were excluded for reasons including undergoing dialysis for chronic renal failure (n=5), recurrent admissions (n=4), and pre-admission blood transfusion (n=1) (Figure 1). Serum copeptin was 2.972 (2.386-3.61) ng/mL among patients with GI bleeding and 2.663 (2.00-3.18) ng/mL among those without GI bleeding (p=0.026) (Table 1).

Of the 118 patients, 75(63.6%) were male; the median age was 72 years (IQR: 58-83). A total of 94 patients (79.7%) had chronic diseases, with hypertension (43%) and diabetes mellitus (31.4%) being the most common (Table 2).

Three patients were discharged after receiving treatment in the ED and refused admission to the ward or intensive care unit, and the median duration of symptoms was 24 (5-24) hours. The median length of hospital stay was found to be 4 (3-5) days. Among, among patients with upper GI bleeding, no correlation was found between serum copeptin levels and the duration of symptoms, length of hospital stay, or number of erythrocyte suspension transfusions in the ED (Table 3).

Within 30 days of diagnosis, 18 (15%) patients died. Serum copeptin levels of the deceased patients did not differ from the survivors (3.28±1.79 ng/mL vs. 3.63±2.76 ng/mL, respectively). At admission, these patients had lower systolic blood pressure and mean arterial pressure and higher shock index values than those who survived. Patients older than 75 years had a higher 30-day mortality rate (p=0.007) (Table 4).

Table 1. Copeptin levels and age and gender characteristics of patients diagnosed with gastrointestinal bleeding

		Gastrointestinal bleeding diagnosis				
		No (n=78)*		Yes (n=118)*		p
		Median (IQR)		Median (IQR)		
Copeptin level (ng/mL)		2.663 (2.00-3.18)		2.972 (2.38-3.61)		**0.026**
Age		69 (55-82)		72 (58-83)		0.335
Gender		N	%	n	%	0.175
	Male	42	53.80	75	63.60	
	Female	36	46.20	43	36.40	

*Of the 196 patients initially presenting with a preliminary diagnosis of gastrointestinal bleeding, the diagnosis was confirmed in 118 patients, whereas it was ruled out in 78 patients. IQR: Interquartile range

Table 2. Distribution of patients' chronic diseases		
	n	%
Presence of chronic disease	94	79.7
Chronic diseases[a]		
Hypertension	51	43.0
Diabetes mellitus	37	31.4
Coronary artery disease	34	28.8
Malignancy	21	17.8
Hepatic disease	12	10.2
Congestive heart failure	11	9.3
Chronic obstructive pulmonary disease	10	8.5
Others (asthma, chronic renal failure, rheumatological disease)	12	9.7
Medication		
Antihypertensives	49	41.50
Antiaggregants	48	40.70
Anticoagulants	38	32.20
Antidiabetics	33	28.00
Proton pump inhibitors	16	13.60
Nonsteroidal anti-inflammatory drugs	10	8.50
[a]There may be more than one chronic disease		

We performed logistic regression analysis to determine the factors affecting mortality and observed that, in patients aged 75 years and older, a shock index above 0.70 [p=0.017, odds ratio (OR)=6.824; 95% confidence interval (CI): (1.419-32.811)] was associated with increased 30-day mortality. Low systolic blood pressure at admission to the ED was associated with increased 30-day mortality [p=0.003, OR=0.952; 95% CI: (0.922-0.983)] (Table 5).

Discussion

In our study, which aimed to demonstrate the effectiveness of copeptin levels in predicting hospitalization duration and 30-day mortality, we observed no association between these predictors. Among markers reported in the literature for assessing prognosis of GI bleeding, we found that the shock index, also considered a vital sign, and age of 75 years or older were better predictors of 30-day mortality than copeptin. Additionally, we found a weak correlation between serum copeptin levels and the amount of FFP transfused.

Table 3. Correlation analysis of serum copeptin levels and symptom duration, length of hospital stay, amount of erythrocyte suspension, and fresh frozen plasma transfusions administered in the ED of patients with upper gastrointestinal bleeding						
		Age	Symptom duration (hours)	Length of stay in hospital (days)	Amount of ES transfusion in the ED	Amount of FFP transfusion in the ED
Copeptin level (ng/mL)	R	-0.069	-0.099	-0.012	-0.012	0.201
	p	0.456	0.287	0.898	0.894	0.261
ED: Emergency department; ES: Erythrocyte suspension; FFP: Fresh frozen plasma						

Table 4. Comparison of age, copeptin levels, vital signs, and laboratory parameters in surviving and deceased patients					
	30 days mortality (mean ± SD)				P
	Survivors (n=100)		Deceased (n=18)		
Copeptin level (ng/mL)	3.631±2.759		3.277±1.786		0.513
Systolic blood pressure (mmHg)	119±23		98±17		**<0.001**
Diastolic blood pressure (mmHg)	66±14		56±9		**0.001**
Mean arterial pressure	83.67±15.65		69.7±10.52		<0.001
Pulse rate (beats/min)	94±20		97±15		0.319
Shock index	0.82±0.25		1.03±0.33		**0.002**
Lactate (mmol/L)	3.0±2.3		3.6±3.8		0.847
BE	4.7±4.5		7.8±6.7		0.108
pH	7.4±0.08		7.38±0.12		0.270
Age	N	%	N	%	
<65	37	37	2	11.1	**0.007**
65≤ - <75	25	25	2	11.1	
75≤	38	38	14	77.8	
BE: Base excess, SD: Standard deviation					

Table 5. Logistic regression analysis of factors associated with 30-day mortality					
Predictors	**P**	**OR**	**95% CI**		**Nagelkerke R^2**
			Lower	**Upper**	
MODEL 1					0.140
Age					
<65	Ref	Ref	Ref	Ref	
65≤ - <75	0.704	1.48	0.195	11.208	
75≤	**0.015**	6.81	1.448	32.087	
MODEL 2					0.230
Age					
<65	Ref	Ref	Ref	Ref	
65≤ - <75	0.642	1.625	0.209	12.62	
75≤	**0.017**	6.824	1.419	32.811	
Shock index					
<0.7	Ref	Ref	Ref	Ref	
≥0.7	**0.045**	8.409	1.045	36.634	
MODEL 3					0.300
Age					
<65	Ref	Ref	Ref	Ref	
65≤ - <75	0.767	1.373	0.169	11.128	
75≤	**0.03**	5.889	1.187	29.215	
Systolic blood pressure	**0.003**	0.952	0.922	0.983	

Backward logistic regression method was used. Since systolic blood pressure and shock index were highly correlated, they were not included in the same model
Model 1: Age is included in the model.
Model 2: Age and shock index are included in the model.
Model 3: Age and systolic blood pressure are included in the model.
OR: Odds ratio, CI: Confidence interval

In studies conducted with healthy volunteers, there is evidence indicating that copeptin levels are influenced by gender, with levels being significantly higher in men (8,14). For instance, Salt et al. (5) found no association between gender and copeptin levels in their study of patients with GI bleeding. Our study revealed no significant correlation between copeptin levels and age, consistent with findings from previous research (7,8). However, we observed an increase in 30-day mortality among individuals aged 75 and older, aligning with reports from earlier studies indicating a higher mortality rate in older individuals (15). Notably, our findings corroborate the literature, which has consistently highlighted age as a significant factor influencing mortality outcomes in patients with GI bleeding. For instance, a study conducted between 1999-2003 on patients with GI bleeding demonstrated varying 90-day mortality rates across different age groups, with mortality rates rising with increasing age (16).

Although copeptin has been studied as a prognostic and severity marker for many medical and traumatic acute conditions (11-13,17,18), our study found no relationship between serum copeptin levels and 30-day mortality in patients admitted to the ED due to GI bleeding. That most of the vital signs of the patients included in our study at admission were within normal ranges suggests that they did not yet have a significant fluid deficit and that the effects of acute stress were not clearly evident. This may be explained by the absence of an association between copeptin and prognosis in patients with GI bleeding, under the aforementioned circumstances. Therefore, future studies may consider collecting serial measurements of serum copeptin and calculating the delta copeptin value in cases of vital sign changes and severe bleeding.

The presumed prognostic value of copeptin may be justified by the positive correlation between individual stress levels and the magnitude of the stressor, or in other words, the severity of the disease (7,19). Copeptin levels were elevated in trauma patients with hemorrhagic shock, predicting the need for massive transfusions (18). In patients with severe bleeding who require massive transfusion, erythrocyte suspensions are supplemented with FFP and platelet transfusions to ensure adequate management of blood loss and coagulopathy (18,20,21). The number of critically ill patients requiring emergency transfusion

in our cohort was limited, and the average vital signs of the patients were within normal ranges. However, a considerable proportion of patients were using anticoagulants. Taken together, these findings suggest that further studies with larger patient populations incorporating bleeding severity classifications and subgroup analyses of patients on anticoagulant therapy are needed to better elucidate the relationship between copeptin levels FFP and ES administration, thereby contributing to the literature and improving patient care processes.

In our investigation, we extended the application of the shock index primarily studied in trauma patients and other hemorrhagic conditions to patients with GI bleeding. In addition to prior studies, which have linked the shock index with thirty-day mortality (22,23), by establishing a cut-off value of 0.7, we found that this threshold correlates with an elevated 30-day mortality risk in patients with GI bleeding. Notably, a previous study reported inferior outcomes for the shock index compared to other prognostic scores regarding 30-day mortality, casting doubt on its clinical utility (24). Consequently, the literature presents conflicting evidence regarding the prognostic value of the shock index. In our study, although patients with fatal outcomes from GI bleeding exhibited high shock index values, the prognostic significance of the shock index was accentuated when considered alongside age. We hypothesize that the limited sample size may have contributed to this observation.

Study Limitations

A primary limitation of our study is the relatively small number of patients. This can be attributed to our city having multiple hospitals equipped to manage GI bleeding cases. Thus, such patients are often transported to other state hospitals via the pre-hospital health care system. Therefore, our dataset may have been insufficient to establish a definitive relationship between serum copeptin levels and prognosis. Additionally, we acknowledge that incorporating a standardized classification of GI bleeding severity could have provided clearer insight into the relationship between copeptin levels and vital signs across different degrees of bleeding. Nevertheless, our study revealed a significant association between the shock index and prognosis. Further research is needed to validate the hypothesis regarding the impact of the shock index on prognosis in the context of GI bleeding.

Conclusion

While serum copeptin levels failed to predict 30-day mortality in patients admitted to the ED for GI bleeding, our model demonstrated that age over 75 years, lower systolic blood pressure, and the shock index, easily calculated from vital signs, were superior predictors of mortality in this patient group.

Ethics

Ethics Committee Approval: The ethical suitability of this research was approved at the meeting of Pamukkale University the Non-Interventional Clinical Research Ethics Committee (decision number: 10, date: 25.05.2021).

Informed Consent: Written informed consent was obtained from the patients to be involved in the study.

Acknowledgments: The data presented in this article are derived from the medical specialization thesis of one of the authors (Emrah Akın).

Footnotes

Authorship Contributions

Concept: E.A., A.O., M.A., M.Ö., O.T.E., M.B., K.İ.T., Design: E.A., A.O., Ö.K.E., M.A., M.Ö., O.T.E., M.B., K.İ.T., Data Collection or Processing: E.A., A.O., Ö.K.E., M.A., M.Ö., A.Y., M.S., O.T.E., M.B., K.İ.T., Analysis or Interpretation: E.A., A.O., Ö.K.E., M.A., M.Ö., A.Y., M.S., M.B., K.İ.T., Literature Search: E.A., A.O., M.A., O.T.E., Writing: E.A., A.O., M.A., A.Y., M.S., O.T.E.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: This work was supported by Pamukkale University, Scientific Research Projects Coordination Unit (Project No: 2021TIPF017).

Use of AI Tools: The authors declare that they did not use artificial intelligence (AI) tools in the creation of this article.

References

1. Ziebell CM, Kitlowski AD, Welch J, Friesen P. Upper gastrointestinal bleeding. In: Tintinalli JE, Ma OJ, Yealy DM, Meckler GD, Stapczynski JS, Cline DM, et al., editors. Tintinalli's emergency medicine: a comprehensive study guide. McGraw-Hill Education. 2020;9:495-8.
2. Lo B. Lower gastrointestinal bleeding. In: Tintinalli JE, Ma OJ, Yealy DM, Meckler GD, Stapczynski JS, Cline DM, et al, editors. Tintinalli's Emergency medicine: a comprehensive study guide. McGraw-Hill Education. 2020;9:498-500.
3. Vora P, Pietila A, Peltonen M, Brobert G, Salomaa V. Thirty-year incidence and mortality trends in upper and lower gastrointestinal bleeding in Finland. JAMA Netw Open. 2020;3:e2020172.
4. Peery AF, Crockett SD, Barritt AS, Dellon ES, Eluri S, Gangarosa LM, et al. Burden of gastrointestinal, liver, and pancreatic diseases in the United States. Gastroenterology. 2015;149:1731-41.e3.
5. Salt O, Durukan P, Ozkan S, Saraymen R, Sen A, Yurci MA. Plasma copeptin levels in the patients with gastrointestinal bleeding. Am J Emerg Med. 2017;35:1440-3.
6. Holwerda DA. A glycopeptide from the posterior lobe of pig pituitaries. I. Isolation and characterization. Eur J Biochem. 1972;28:334-9.

7. Koch A, Yagmur E, Hoss A, Buendgens L, Herbers U, Weiskirchen R, et al. Clinical relevance of copeptin plasma levels as a biomarker of disease severity and mortality in critically ill patients. J Clin Lab Anal. 2018;32:e22614.
8. Morgenthaler NG, Struck J, Alonso C, Bergmann A. Assay for the measurement of copeptin, a stable peptide derived from the precursor of vasopressin. Clin Chem. 2006;52:112-9.
9. Morgenthaler NG, Struck J, Jochberger S, Dünser MW. Copeptin: clinical use of a new biomarker. Trends Endocrinol Metab. 2008;19:43-9.
10. Darzy KH, Dixit KC, Shalet SM, Morgenthaler NG, Brabant G. Circadian secretion pattern of copeptin, the C-terminal vasopressin precursor fragment. Clin Chem. 2010;56:1190-1.
11. Abdelaziz AA, Khattab AA, Abdelmaksoud MH, Ghazy RM, Noaman A. Plasma copeptin as a prognostic marker in children with heart failure. Indian Pediatr. 2024;61:1103-8.
12. Aguilera G, Subburaju S, Young S, Chen J. The parvocellular vasopressinergic system and responsiveness of the hypothalamic pituitary adrenal axis during chronic stress. Prog Brain Res. 2008;170:29-39.
13. Choi KS, Kim HJ, Chun HJ, Kim JM, Yi HJ, Cheong JH, et al. Prognostic role of copeptin after stroke: a systematic review and meta-analysis of observational studies. Sci Rep. 2015;5:11665.
14. Bhandari SS, Loke I, Davies JE, Squire IB, Struck J, Ng LL. Gender and renal function influence plasma levels of copeptin in healthy individuals. Clin Sci (Lond). 2009;116:257-63.
15. Benedeto-Stojanov D, Bjelaković M, Stojanov D, Aleksovski B. Prediction of in-hospital mortality after acute upper gastrointestinal bleeding: cross-validation of several risk scoring systems. J Int Med Res. 2022;50:3000605221086442.
16. Wilcox CM, Cryer BL, Henk HJ, Zarotsky V, Zlateva G. Mortality associated with gastrointestinal bleeding events: Comparing short-term clinical outcomes of patients hospitalized for upper GI bleeding and acute myocardial infarction in a US managed care setting. Clin Exp Gastroenterol. 2009;2:21-30
17. Hsein YC, Wu IJ, Tan J, Huang SS, Lu KT, Su CH, et al. Serum levels of copeptin predict adverse outcomes and improve risk prediction of TRISS and MGAP scores in patients with multiple trauma: a single-center prospective cohort study. J Trauma Acute Care Surg. 2023;94:336-43.
18. Sims CA, Guan Y, Bergey M, Jaffe R, Holmes-Maguire L, Martin N, et al. Arginine vasopressin, copeptin, and the development of relative AVP deficiency in hemorrhagic shock. Am J Surg. 2017;214:589-95.
19. Katan M, Fluri F, Morgenthaler NG, Schuetz P, Zweifel C, Bingisser R, et al. Copeptin: a novel, independent prognostic marker in patients with ischemic stroke. Ann Neurol. 2009;66:799-808.
20. Yang DB, Yu WH, Dong XQ, Du Q, Shen YF, Zhang ZY, et al. Plasma copeptin level predicts acute traumatic coagulopathy and progressive hemorrhagic injury after traumatic brain injury. Peptides. 2014;58:26-9.
21. Meißner A, Schlenke P. Massive bleeding and massive transfusion. Transfus Med Hemother. 2012;39:73-84.
22. Dogru U, Yuksel M, Ay MO, Kaya H, Ozdemır A, Isler Y, et al. The effect of the shock index and scoring systems for predicting mortality among geriatric patients with upper gastrointestinal bleeding: a prospective cohort study. Sao Paulo Med J. 2022;140:531-9.
23. Lowe J, Lowe Z, Ely R. Shock index as a predictor of mortality and hospital admission in prehospital gastrointestinal bleeding: a retrospective cohort study. Prehosp Emerg Care. 2024;28:689-95.
24. Saffouri E, Blackwell C, Laursen SB, Laine L, Dalton HR, Ngu J, et al. The shock index is not accurate at predicting outcomes in patients with upper gastrointestinal bleeding. Aliment Pharmacol Ther. 2020;51:253-60.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.43650
Eurasian J Emerg Med. 2026;25: 315-9

Analysis of the Prognostic Value of Carboxyhemoglobin in Patients with Acute Coronary Syndrome

Alpaslan Ünlü[1], Zeynep Çakır[2]

[1]Yalova Provincial Health Directorate, Yalova Training and Research Hospital, Clinic of Emergency Medicine, Yalova, Türkiye
[2]Atatürk University Faculty of Medicine, Department of Emergency Medicine, Erzurum, Türkiye

Abstract

Aim: While carbon monoxide is a gas that is toxic when introduced exogenously into the body, it also functions as a signaling molecule with regulatory roles in many physiological processes within the cardiovascular system. The aim of this study is to investigate the utility of endogenous carboxyhemoglobin (COHb) for estimating mortality within the first three months after admission among patients with acute coronary syndrome.

Materials and Methods: Our study was conducted on patients who came to the emergency service with a complaint of chest pain, met the inclusion criteria, were hospitalized in the cardiology clinic. Patients' demographic characteristics, coronary angiography results, electrocardiogram, blood gas assessment, and blood tests' prognosis were recorded.

Results: The average age was 61.52±13.19 (min: 21 - max: 93) and 164 (71.9%) patients were male. Angiography results showed that 110 (48.2%) patients had obstruction of the left anterior descending artery, and 82 (36%) patients had obstruction of the right coronary artery. In the analysis conducted after excluding patients who had been exposed to cigarette smoke before hospital admission, endogenous COHb level was not a statistically significant predictor of in-hospital mortality (p=0.248) or three-month mortality (p=0.26).

Conclusion: No statistically significant relationship was found between endogenous COHb levels and predicted mortality in patients diagnosed with acute coronary syndrome who were neither exposed to cigarette smoke nor had a history of smoking before hospital admission. However, we think that the small patient population, together with limitations in the study may have limited the findings.

Keywords: Carbon monoxide, acute coronary syndrome, myocardial infarction, unstable angina, acute coronary syndrome prognosis

Introduction

It has been predicted that carbon monoxide (CO), which has a toxic effect, also acts as signal molecule which has a regulatory role in many pathophysiological processes that take place in the cardiovascular, immune and nervous system (1,2).This effect results from the modulation of soluble guanylate cyclase and upregulation of cyclic guanosine monophosphate production similar to nitric oxide (3). In addition, studies which show its anti-apoptotic, anti-inflammatory and vasodilatation effects have been conducted (2,4,5). On the other hand, endogenous production increases in a number of haematological diseases caused by the destruction of hemoproteins such as haemolytic anemia, hematoma, thalassemia and Gilbert's syndrome (6,7). It has also been found that endogenous CO production increases before treatment in critical diseases such as sepsis, septic shock and cirrhosis (8,9). Although carboxyhemoglobin (COHb) levels are high in healthy smoking adults, endogenous CO level increases in different inflammatory and oxidative pathological conditions (10). This study examines the usability of endogenous COHb in the prediction of mortality in the first three months after admission to hospital in acute coronary syndrome (ACS) patients (11).

Materials and Methods

Study Protocol

This study was planned prospectively and included 228 patients older than 18 years of age who presented to the

Corresponding Author: Alpaslan Ünlü MD, Yalova Provincial Health Directorate, Yalova Training and Research Hospital, Clinic of Emergency Medicine, Yalova, Türkiye
E-mail: alpiunlu@gmail.com **ORCID ID:** orcid.org/0000-0001-6427-4594

Cite this article as: Ünlü A, Çakır Z. Analysis of the prognostic value of carboxyhemoglobin in patients with acute coronary syndrome. Eurasian J Emerg Med. 2026;25: 315-9.

Received: 26.11.2025
Accepted: 29.04.2026
Published: 13.05.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Atatürk University Faculty of Medicine Emergency Clinic with symptoms suggestive of acute coronary artery disease, such as chest pain, shortness of breath, tachycardia, and back pain, and who were hospitalized in the cardiology clinic. On admission, in addition to an ACS assessment, a blood gas assessment was performed (Radiometer ABL 800-flex-Radiometer Medical ApS, Bronshoj, Denmark). In addition, the patient's demographic and clinical characteristics such as age and gender, exposure to cigarette smoke, past medical history, medications used, vital signs, coronary angiography results, echocardiography results, hospitalization duration, and prognosis were recorded. Patients who did not volunteer to participate in the study; patients with comorbid conditions that could affect blood CO levels; patients whose vital signs were unstable; patients who had cardiopulmonary arrest; patients who received respiratory support; patients with known chronic lung disease and patients who had smoked or been exposed to cigarette smoke immediately before arriving at the emergency service were excluded from the study (Figure 1). In addition, patients with incomplete echocardiographic or coronary angiographic data due to early discharge, referral to another center, or incomplete in-hospital evaluation were excluded from the analysis. The study protocol was approved by the Atatürk University Faculty of Medicine Ethical Board (ethical board number: B.30.2.ATA.0.01.00/, date: 08.12.2016).

Hemodynamic instability was defined as any of the following: systolic blood pressure <90 mmHg, requirement for vasopressor support, need for invasive or non-invasive mechanical ventilation, or cardiopulmonary arrest at admission. These patients were excluded to avoid the confounding effects of hypoxia and exogenous oxygen therapy on COHb measurements.

Statistical Analysis

For statistical analyses, the data were recorded with the Statistical Package for Social Sciences (SPSS 20) program. A Paired-sample t test was used to compare dependent variables that were normally distributed. All data were expressed as average ± standard deviation and $p<0.05$ level was considered as statistically significant.

To test the predefined study hypothesis, a minimum sample size of 128 patients was calculated to detect an effect size of 0.80 with a statistical power of 80% and a type I error rate of 0.05. A total of 228 patients were included in the final analysis, exceeding the required sample size.

Results

During the study period, 736 patients with a diagnosis of coronary artery disease were admitted to the cardiology clinic from the emergency department. One hundred forty-nine patients declined to participate in the study. Patients with chronic lung disease (n=120); those with a history of smoking or exposure to cigarette smoke before admission (n=160); those referred to the emergency service already intubated or receiving respiratory support with unstable hemodynamics (n=66) and patients for whom echocardiography and angiography data could not be retrieved from the hospital automation system

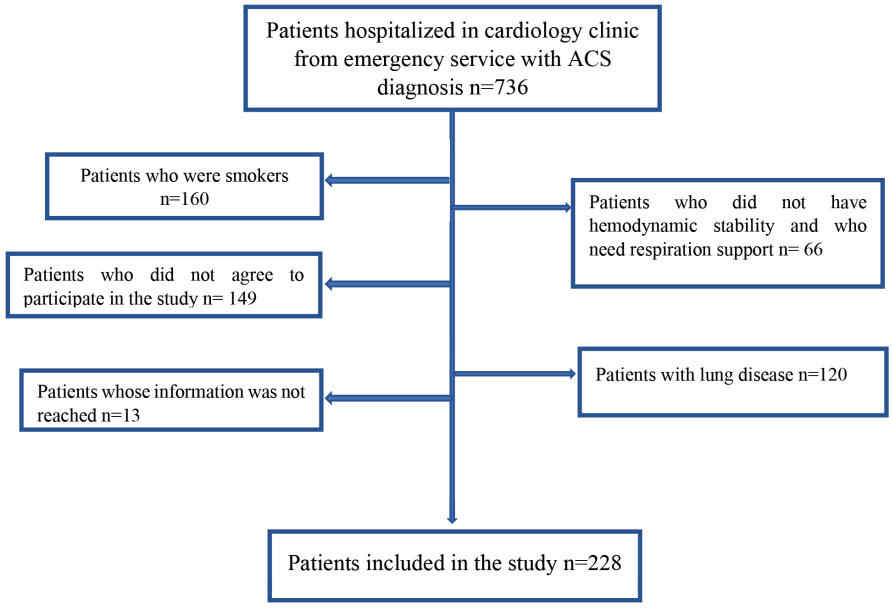


Figure 1. Flow chart of the patients included in the study
ACS: Acute coronary syndrome

(n=13) were excluded from the study, resulting in a final sample of 228 patients (Figure 1).

Of the patients included in the study, 164 (71.9%) were male and average age was 61.52±13.19 (21-93). When the patients' medical histories were examined, 192 (84.2%) had a history of hyperlipidemia, 91 (39.9%) had a history of hypertension, and 71 (31.1%) had a diagnosis of diabetes. When the history of drug use for existing comorbid diseases was examined, 42% (n=96) used aspirin, 25% (n=57) used a beta-blocker, and 15.8% (n=36) used an antilipidemic (Table 1).

When the admission complaints of the patients who were thought to have ACS were examined, 215 (94.5%) had chest pain; on taking anamnesis, 85 (37.3%) patients reported circumstances that increased angina symptoms.

On echocardiography performed after hospital admission, 69 (30.3%) of patients had an ejection fraction below 40%. No statistical association was found between ejection fraction and COHb level (p=0.521). In our study, wall motion disorder was found in 45 (19.7%) patients and no statistically significant association was found between wall motion disorder and endogenous COHb (p=0.172)

When the patients' ECGs were examined, depression was present in 17.1% (n=39), elevation in 36.8% (n=84), and T negativity was present in 30.3% (n=69). 82 (36%) patients with ST elevation myocardial infarction (STEMI) and 112 (49.1%) patients with non-STEMI were included in the study. Angiography was performed in 194 (85.1%) patients who were treated as inpatients for ACS. 110 (48.2%) patients had left anterior descending artery (LAD) obstruction, and 82 (36%) had right coronary artery (RCA) obstruction. No statistical association was found between COHb levels (RCA: p=0.797, LAD: p=0.826, circumflex: p=0.411) and the presence of lesions in any coronary artery among patients who received angiography (p=0.293). Angiography revealed 157 (68.9%) patients with coronary artery obstruction higher than 50%; no statistically significant association was found between the degree of obstruction and endogenous COHb (p=0.498).

Table 1. Demographic characteristics of the patients hospitalized in cardiology clinic from the emergency service

Characteristics	
Age	61.52±13.19 (min: 21-max: 93)
Gender	
Male	164 (71.9%)
Comorbid diseases in the anamnesis	
Hyperlipidemia	192 (84.2%)
Hypertension	91 (39.9%)
Diabetes	71 (31.1%)
By-pass operation	37 (16.2%)
Congestive heart failure	33 (14.5%)
Cerebrovascular disease	20 (8.8%)
Renal failure	14 (6.1%)
Medications used	
Aspirin	96 (42%)
Beta blocker	57 (25%)
Antihyperlipidemic	36 (15.8%)
ACE* inhibitor-ARB**	34 (14.9%)
Clopidegrol	34 (14.9%)
Calcium canal blocker	22 (9.6%)
Admission complaint	
Chest pain	215 (94.5%)
Dyspnea	24 (10.5%)

*ACE: Angiotensin-converting enzyme, **ARB: Angiotensin receptor blocker

When patients' treatments were examined, it was found that 225 (98.7%) patients received medical treatment, 147 (64.5%) patients received percutaneous intervention, 104 (45.6%) patients received a stent, and 8 (3.5%) patients underwent bypass operation.

When the patients' mortality rates were examined, 16 (7%) died during hospitalization, 3 (1.3%) died within a month, and 4 (1.8%) died within three months. Endogenous COHb level was not a statistically significant predictor of in-hospital mortality, one-month mortality (p=0.248), or three-month mortality (p=0.26).

Discussion

Endogenous CO is increasingly recognized as a biologically active molecule with cytoprotective, anti-inflammatory, and vasodilatory properties, primarily mediated through the heme oxygenase (HO) pathway. Experimental and clinical studies have demonstrated that HO-1 induction and subsequent CO production may increase in response to ischemic and oxidative stress. Therefore, COHb has been proposed as a potential surrogate marker reflecting endogenous CO activity and systemic stress responses (2,12,13).

Although endogenous CO production increased in case of ischemic stress (12), we concluded that it was not effective in predicting short-term mortality in patients diagnosed with ACS. However, in a study conducted by Kobayashi et al. (13), it was concluded that endogenous CO can be useful in evaluating cardiovascular stress risk. We believe that this difference between the two studies results from the fact that while increasing endogenous production by causing free hem release from hem-increasing proteins such as haemoglobin, myoglobin and cytochrome in case of ischemic stress, it is not an effective method in determining mortality (12). This discrepancy may be explained by several factors. First, ACS-related mortality is influenced by a multifactorial interplay among clinical severity, extent of myocardial injury, hemodynamic status, and established prognostic markers, such as left ventricular systolic

function, coronary anatomy, and biochemical markers. In this complex clinical setting, the isolated contribution of COHb may be insufficient to independently predict mortality (11,14).

While CO has a toxic characteristic disrupting cellular respiration, it is formed endogenously as a result of the breakdown of hem through HO enzyme and through photooxidation, lipid peroxidation and xenobiotic metabolism (15). It is known that CO, which is synthesized naturally in our body, has physiological functions such as anti-inflammatory, anti-oxidative, anti-apoptotic, vasodilation, angiogenesis and vascular remodeling in low concentrations (16,17).

It has been shown in experimental animals that blood COHb levels can be used to predict HO activity (15). In addition, it has been associated with increased HO-1 expression and elevated COHb levels. It is thought that increased HO-1 expression and high COHb level is associated with more serious course in patients who have critical disease, chronic obstructive pulmonary disease, systemic inflammatory response syndrome, acute respiratory distress syndrome and acute ischemic disease (16,18). One notable example is the retrospective study by Kakavas et al. (18), which examined the relationship between COHb level and prognosis in patients with pulmonary embolism. Univariate logistic regression analysis showed that COHb and methemoglobin levels were associated with mortality, while multivariate analysis showed that COHb was an independent predictor of in-hospital mortality. In our study, no data were found indicating that it is effective in evaluating in-hospital mortality in ACS patients. In patients diagnosed with pulmonary embolism. This can result from a number of chronic clinical situations, such as previous deep vein thrombosis, malignancy, immobilization, and bleeding disorders. We believe that the significant result reported in Kakavas et al.'s study (18) may be attributable to chronic diseases affecting COHb levels.

A large number of studies have been conducted to find out in-hospital and short term mortality in patients diagnosed with ACS (19). In studies conducted, it has been found that the relationship between type B natriuretic peptide (20), platelet neutrophil ratio (21), uric acid (22) and angiographic scoring system and serum albumin level (23) is effective in finding out mortality (24,25,26). However, in our study, endogenous CO was not significantly associated with mortality.

Study Limitations

The present study is a cross-sectional analysis conducted at a single institution. The number of subjects was relatively low. Therefore, prospective studies with larger sample sizes are needed.

Because the number of mortality events was relatively low, multivariate regression analysis could not be performed. Another important limitation of this study is the lack of detailed subgroup-level COHb distribution data for survivors and non-survivors at different follow-up time points. Although mortality outcomes were recorded, the available dataset did not allow a reliable post hoc comparison of COHb levels between survivors and non-survivors. Therefore, such analyses could not be performed and should be addressed in prospective studies that include predefined subgroup analyses.

Conclusion

There is no globally accepted mortality marker that can be used in daily practice for patients with ACS. The results of our study showed that endogenous COHb was not effective in predicting in-hospital and three-month mortality in patients with ACS. Although no statistically significant association was shown in our study, we think that the efficacy of endogenous COHb can be demonstrated more accurately with a larger number of patients and in multicultural studies.

Ethics

Ethics Committee Approval: The study protocol was approved by the Atatürk University Faculty of Medicine Ethical Board (ethical board number: B.30.2.ATA.0.01.00/, date: 08.12.2016).

Informed Consent: This is prospectively study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: A.Ü., Z.Ç., Concept: A.Ü., Z.Ç., Design: A.Ü., Z.Ç., Data Collection or Processing: A.Ü., Z.Ç., Analysis or Interpretation: A.Ü., Z.Ç., Literature Search: A.Ü., Z.Ç., Writing: A.Ü., Z.Ç.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Wu L, Wang R. Carbon monoxide: endogenous production, physiological functions, and pharmacological applications. Pharmacol Rev 2005;57:585-630.
2. Padda I, Sethi Y, Das M, Fabian D, Ralhan T, Aziz D, et al. Heme oxygenase-1, cardiac senescence, and myocardial infarction: a critical review of the triptych. Cardiovasc Drugs Ther. 2025;39:1453-64.
3. Furchgott RF, Jothianandan D. Endothelium-dependent and-independent vasodilation involving cyclic GMP: relaxation induced by nitric oxide, carbon monoxide and light. Blood Vessels 1991;28:52-61.
4. Stein AB, Bolli R, Dawn B, Sanganalmath SK, Zhu Y, Wang OL, et al. Carbon monoxide induces a late preconditioing-mimetic cardioprotective and antiapoptotic milieu in the myocardium. J Mol Cell Cardiol. 2012;52: 228-36.

5. Kobayashi A, İshikawa K, Matsumoto H, Kimura S, Kamiyama Y, Maruyama Y. Synergetic antioxidant and vasodilatory action of carbon monoxide in angiotensinII- induced cardiac hypertrophy. Hypertension. 2007;50:1040-8.
6. James EB, Vreman HJ, Wong RJ, Stevenson DK, Vichinsky E, Schumacher L, et al. Elevated exhaled carbon monoxide concentration in hemoglobinopathies and its relation to red blood cell transfusion therapy. Pediatr Hematol Oncol. 2010;27:112-21.
7. Hampson NB, Weaver LK. Carbon monoxide poisoning: a new incidence for an old disease. Undersea Hyperb Med. 2007;34:163-8.
8. Zegdi R, Perrin D, Burdin M, Boiteau R, Tenailon A. Increased endogenous carbon monoxide production in severe sepsis. Intensive Care Med. 2002;28:793-6.
9. Tarquini R, Masini E, La Villa G, Barietta G, Noveli M, Mastroianni R, Romanelli RG, Vizzutti F, Santosuosso U, Laffi G. Increased plasma carbon monoxide in patients with viral cirrhosis and hyperdynamic circulation. Am J Gastroenterol. 2009;104:891-7.
10. Owens EO. Endogenous carbon monoxide production in disease. Clin Biochem. 2010;43:1183-8.
11. Rao SV, O'Donoghue ML, Ruel M, Rab T, Tamis-Holland JE, Alexander JH, et al. 2025 ACC/AHA/ACEP/NAEMSP/SCAI guideline for the management of patients with acute coronary syndromes: a report of the american college of cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 2025;151:e771-e862. Epub 2025 Feb 27.
12. Zeynalov E, Shah ZA, Li RC, Dore S. Heme oxygenase 1 is associated with ischemic preconditioning-induced protection against brain ischemia. Neurobiol Dis. 2009;35:264-9.
13. Kobayashi A, Mizukami H, Sakamoto N, Yamaki T, Kunii H, Nakazato K, et al. Endogenous carbon monoxide concentration in blood elevates in acute coronary syndrome of nonsmoker population. Fukushima J Med Sci. 2015;61:72-8.
14. Kraler S, Mueller C, Libby P, Bhatt DL. Acute coronary syndromes: mechanisms, challenges, and new opportunities. Eur Heart J. 2025;46:2866-89.
15. Wang QM, Yin XY, Duan ZJ, Guo SB, Sun XY. Role of the heme oxygenase/carbon monoxide pathway in the pathogenesis and prevention of hepatic encephalopathy. Mol Med Rep. 2013;8:67-74.
16. Fazekas AS, Wewalka M, Zauner C, Funk GC. Carboxyhemoglobin levels in medical intensive care patients: a retrospective, observational study. Crit Care. 2012;16:R6.
17. Godai K, Hasegawa-Moriyama M, Kuniyoshi T, Matsunaga A, Kanmura Y. Increased carboxyhemoglobin level during liver resection with inflow occlusion. J Anesth. 2013;27:306-8.
18. Kakavas S, Papanikolaou A, Ballis E, Tatsis N, Goga C, Tatsis G. Carboxyhemoglobin and methemoglobin levels as prognostic markers in acute pulmonary embolism. Am J Emerg Med. 2015;33:563-8.
19. Abubakar M, Irfan U, Abdelkhalek A, Javed I, Khokhar MI, Shakil F, et al. Comprehensive quality analysis of conventional and novel biomarkers in diagnosing and predicting prognosis of coronary artery disease, acute coronary syndrome, and heart failure, a comprehensive literature review. J Cardiovasc Transl Res. 2024;17:1258-85.
20. Jaberg L, Toggweiler S, Puck M, Frank M, Rufibach K, Luscher TF, et al. Prognostic value of N-terminal pro-B-type natriuretic peptide in patients with acute coronary syndromes undergoing left main percutaneous coronary intervention. Circ J. 2011;75: 2648-53.
21. Li H, Zhou Y, Ma Y, Han S, Zhou L. The prognostic value of the platelet-tolymphocyte ratio in acute coronary syndrome: a systematic review and metaanalysis. Kardiol Pol. 2017;75:666-73.
22. Magnoni M, Berteotti M, Ceriotti F, Mallia V, Vergani V, Peretto G, et al. Serum uric acid on admission predicts in-hospital mortality in patients with acute coronary syndrome. Int J Cardiol. 2017;240:25-9.
23. Kurtul A, Murat SN, Yarlioglues M, Duran M, Ocek AH, Koseoglu C, et al. Usefulness of serum albumin concentration to predict high coronary SYNTAX score and in-hospital mortality in patients with acute coronary syndrome. Angiology. 2016;67:34-40.
24. Pruc M, Kubica J, Banach M, Swieczkowski D, Rafique Z, Peacock WF, et al. Diagnostic and prognostic performance of the neutrophil-to-lymphocyte ratio in acute coronary syndromes: a meta-analysis of 90 studies including 45 990 patients. Kardiol Pol. 2024;82:276-84.
25. Pruc M, Kubica J, Banach M, Swieczkowski D, Rafique Z, Peacock WF, et al. Prognostic value of the monocyte-to-high-density lipoprotein-cholesterol ratio in acute coronary syndrome patients: a systematic review and meta-analysis. Kardiol Pol. 2025;83:52-61.
26. Yildirim M, Salbach C, Mueller-Hennessen M, Frey N, Giannitsis E. Diagnostic and prognostic evaluation of novel biomarkers compared to ESC 0/1 h and 0/3 h algorithms in patients with suspected Non-ST-elevation myocardial infarction. J Clin Med. 2025;14.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.02525
Eurasian J Emerg Med. 2026;25: 320-8

Association between Anxiety Levels of Patient Relatives and Violence Tendencies in the Emergency Department

Seda Öztürk, Durdane Yılmaz Güven

Karabük University, Graduate Education Institute, Nursing Science, Karabük, Türkiye

Abstract

Aim: To examine the association between anxiety, depression, and tendency toward violence among relatives of patients in emergency departments.

Materials and Methods: The study was conducted among 371 relatives of patients in the emergency department of a district state hospital between February and May 2025. Data were collected using a personal information form the Hospital Anxiety and Depression scale (HADS), and the Violence Tendency scale (VTS).

Results: The mean HADS-anxiety score of patient relatives was 8.16±3.93, and the mean depression score was 7.10±3.66. The mean VTS score was 36.97±9.19. Analysis showed a positive, weak, and statistically significant association between the VTS and the HADS-anxiety subscale ($r=0.254$; $p<0.001$), and between the VTS and the HADS-depression subscale ($r=0.218$; $p<0.001$). Multiple linear regression analysis showed that anxiety and depression scores were significantly associated with a tendency toward violence.

Conclusion: Anxiety and depression levels among relatives of patients were significantly associated with their tendencies toward violence. Emergency department staff and administrators should develop strategies to prevent violence and to reduce the anxiety of patients' relatives.

Keywords: Emergency department, patient relatives, anxiety, violence tendency

Introduction

Emergency departments, which are the busiest, most stressful, and most time-pressured units in healthcare systems, are environments of high emotional and psychological strain for both patients and their relatives. Within this dynamic structure, nurses are the healthcare professionals who communicate most with patients and their relatives, serve as the first point of contact, and play a critical role in crisis management (1). Relatives of patients visiting emergency departments experience uncertainty regarding the lives of their loved ones, which can lead to intense emotions such as anxiety, stress, and fear (2). Traumatic cases and critical health conditions, in particular, can challenge the psychological resilience of patient relatives and increase their levels of anxiety (3).

Anxiety is a psychological response to perceived threats, which can reduce rational thinking and lead to reactive behaviors (4). The literature indicates that individuals with high levels of anxiety may exhibit violence tendencies in social interactions, and in environments where a sense of control is lost, these behaviors can escalate into violence (5). The emergency department is a high-risk psychosocial setting for both patient relatives and healthcare professionals due to the intensity of uncertainties, limited information sharing, and heavy workload (6,7).

Research conducted in Türkiye and worldwide indicates that factors such as lack of information, insufficient communication, inadequate physical conditions, and prolonged waiting times in emergency departments can trigger violent behaviors by patient relatives toward healthcare professionals, particularly nurses who often serve as the first point of contact (8-11). Violence against healthcare workers not only threatens staff safety but also negatively affects the quality of healthcare delivery (12). Among the psychological factors contributing to such violent behaviors, anxiety has been identified as a significant determinant (13).

Corresponding Author: Durdane Yılmaz Güven MD, Karabük University, Graduate Education Institute, Nursing Science, Karabük, Türkiye
E-mail: durdaneguven@karabuk.edu.tr **ORCID ID:** orcid.org/0000-0001-8044-1014

Cite this article as: Öztürk S, Yılmaz Güven D. Association between anxiety levels of patient relatives and violence tendencies in the emergency department. Eurasian J Emerg Med.2026;25: 320-8.

Received: 24.11.2025
Accepted: 24.05.2026
Published: 11.06.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

In emergency settings, deficiencies in information provision, ineffective communication, and the inability of healthcare personnel to allocate adequate time to patient relatives due to heavy workloads may increase anxiety levels and contribute to conflictual behaviors and violence tendencies (14). Consequently, this study was conducted to determine the relationship between anxiety and depression levels in relatives of patients and their tendencies toward violence.

Materials and Methods

Research Design

This study used an analytical cross-sectional (correlational) design, a quantitative research method.

Study Setting and Time

The research was conducted between February and May 2025 in the emergency department of a secondary-level district state hospital. The emergency department receives an average of 650 patient visits per day, indicating a high patient turnover. Data were collected during peak hours when family members of patients meeting the inclusion criteria were present in the emergency department.

Population and Sample

The study sample consisted of relatives of patients treated in the emergency department of the district state hospital. The sample was obtained by inviting relatives of patients treated in the yellow and green zones of the emergency department and selecting them using simple random sampling. Individuals in the sample were over 18 years of age, had no communication difficulties, were fluent in Turkish, and voluntarily agreed to participate in the study.

To justify the sample size, a post-hoc power analysis was performed using G*Power 3.1.9.7 software. For the multiple linear regression model involving two predictors (anxiety and depression) and a sample size of 371, the statistical power was calculated to be greater than 0.95 at an alpha level of 0.05 for a small-to-medium effect size (f^2=0.08). This result confirms that the study has sufficient power to detect the reported associations between psychological factors and the tendency toward violence.

Data Collection

Participants who agreed to take part in the study were provided with written and verbal information about the study's purpose, confidentiality, and voluntary participation, and they signed informed consent forms. The personal information form, the Hospital Anxiety and Depression scale (HADS), and the Violence Tendency scale (VTS) were then administered face-to-face by the researcher. The personal information form developed by the researcher, was used to determine the demographic characteristics of the participants, such as age, gender, education level, marital status, and frequency of emergency department visits. The HADS, developed by Zigmond and Snaith (15) and validated in Türkiye by Aydemir et al. (16), is designed to assess anxiety and depression levels in hospitalized individuals. It consists of 14 items, including 7 assessing anxiety and 7 assessing depression. In this study, both the anxiety and depression subscales were used. Scores are evaluated using a 4-point Likert scale, with higher scores indicating higher levels of anxiety. In Aydemir et al. (16) study, the Cronbach's alpha value was 0.8525 for the anxiety subscale and 0.7784 for the depression subscale. In the current study, the Cronbach's alpha was 0.744 for the anxiety subscale and 0.649 for the depression subscale. The VTS was developed by the Turkish Prime Ministry Family Research Institution (17). This 20-item scale uses a 4-point Likert format to measure individuals' attitudes toward aggression and violent behaviors. Scores range from 1 to 80, with higher scores indicating a higher tendency toward violence. This scale is considered an efficient tool for assessing violence tendencies in individuals (17). In its reliability study, Cronbach's alpha 0.87; in the current study it was 0.854. Data collection took approximately 15-20 minutes per participant, with explanatory support provided as needed.

Prior to the study, all necessary legal and ethical approvals were obtained. The research protocol was reviewed and approved by Karabük University Rectorate Social and Human Sciences Research Ethics Committee (decision number: 2025/01, date: 03.01.2025), and institutional permission was obtained from the district state hospital where the study was conducted. Participants were informed verbally and in writing about the purpose, scope, data collection process, and voluntary nature of the study. Participation was entirely voluntary, and participants were informed of their right to withdraw from the study at any time without providing a reason. Informed consent was obtained from all participants.

Throughout the study, strict adherence to participant confidentiality and data security principles was maintained. Personal information was not shared with third parties; data were used solely for scientific purposes, and anonymity was preserved. The study was conducted in accordance with the ethical principles outlined in the Helsinki Declaration.

Statistical Analysis

The data were analyzed using the IBM SPSS Statistics for Windows, version 25.0 (IBM Corp., Armonk, NY, USA). The data obtained from the study were analyzed using descriptive statistics. Categorical variables are presented in summary form using frequency (n) and

percentage (%) values. Descriptive statistics, Pearson correlation, multiple linear regression, independent- samples t-test, and ANOVA were performed. The level of statistical significance was set at $p<0.05$. Pearson correlation analysis was performed to examine the relationships among anxiety, depression, and tendency toward violence. To investigate whether anxiety and violence tendency scores differed according to demographic variables, appropriate analyses such as independent samples t-test, one-way ANOVA, and, if necessary, chi-square tests were applied (18). A significance level of $p<0.05$ was accepted.

Results

A total of 371 relatives of patients participated in the study. Examination of the participants' age distribution showed that the largest group was aged 18–25 (n=157, 42.3%). 69.5% of the participants were women (n=258), and 61.4% (n=228) were university graduates. Of the participants, 55.5% (n=206) were single and 38.0% (n=141) were employed. When health habits were evaluated, 62.8% of the participants had no specific habits (n=233). Among individuals reporting a habit, smoking was the most common behavior (n=112, 30.2%) (Table 1).

Table 1. Socio-demographic characteristics of the participants

Variable	Category	n (%)
Age	18-25 years	157 (42.3)
	26-35 years	83 (22.4)
	36-45 years	83 (22.4)
	46-55 years	41 (11.0)
	56 years and above	7 (1.9)
Gender	Female	258 (69.5)
	Male	113 (30.5)
Education	University	228 (61.4)
	High school	63 (17.0)
	Master's/Doctorate	45 (12.1)
	Primary school	24 (6.5)
	Middle school	11 (3.0)
Marital status	Married	165 (44.5)
	Single	206 (55.5)
Occupation	Employed	141 (38.0)
	Student	128 (34.5)
	Housewife	56 (15.1)
	Other	32 (8.6)
	Retired	14 (3.8)
Habits	No habits	233 (62.8)
	Smoking	112 (30.2)
	Alcohol	26 (7.0)

The vast majority of participants had previously visited a healthcare institution (n=341, 91.9%), and according to the distribution of diagnoses, the most common disease groups were gastrointestinal system diseases (n=126, 34.0%) and respiratory system diseases (n=113, 30.5%). Regarding treatment outcomes, 72.0% (n=267) of participants reported a positive improvement in their patients, while the highest level of anxiety was observed in the "slightly anxious" group (n=202, 54.5%). Regarding their history of violence, almost all participants reported having never been involved in any violent incident before (n=368, 99.2%). The situation that angered or disturbed participants the most was "not being paid attention to" (n=111, 29.9%). This was followed by a lack of patient information and a crowded hospital environment (both n=81, 21.8%). When redress behavior was examined, the highest percentage of participants indicated they would file a formal complaint (n=181, 48.8%) (Table 2).

Table 2. Participants' experiences, emotional states, and hospital-related characteristics

Variable	Category	n (%)
Psychiatric disorder	No	333 (89.8)
	Yes	38 (10.2)
Relationship with patient	Family	187 (50.4)
	Friend	96 (25.9)
	Other	88 (23.7)
Previous hospital visit	Yes	341 (91.9)
	No	30 (8.1)
Type of illness	Gastrointestinal system diseases	126 (34.0)
	Respiratory system diseases	113 (30.5)
	Musculoskeletal system diseases	46 (12.4)
	Neurological diseases	43 (11.6)
	Trauma	19 (5.1)
	Urinary system diseases	12 (3.2)
	Cardiac system diseases	12 (3.2)
Treatment outcome	Positive improvement	267 (72.0)
	No improvement	104 (28.0)
Level of concern	Slightly concerned	202 (54.5)
	Not concerned at all	120 (32.3)
	Moderately concerned	33 (8.9)
	Very concerned	16 (4.3)

Table 2. Continued		
Variable	**Category**	**n (%)**
History of violence	No	368 (99.2)
	Yes	3 (0.8)
Most upsetting/angering reasons	Not being addressed	111 (29.9)
	Lack of information about patient	81 (21.8)
	Crowded hospital environment	81 (21.8)
	Long treatment duration	31 (8.4)
	Other	27 (7.3)
	Insufficient hospital comfort	13 (3.5)
	No available bed for patient	12 (3.2)
	Healthcare staff shouting	9 (2.5)
	Patient's condition not improving	6 (1.6)
Seeking justice	I will file a formal complaint	181 (48.8)
	Do not feel victimized	83 (22.4)
	Manage/I do nothing	51 (13.7)
	No response	38 (10.2)
	Other	18 (4.9)

In the study, the mean HADS-anxiety score of patients' relatives was 8.16±3.93, indicating a mild level of anxiety; the mean HADS-depression score was 7.10±3.66, indicating a low/normal level of depression; and the mean VTS score was 36.97±9.19, indicating a low level of violence tendency (Table 3).

As shown in Table 3, there were statistically significant relationships between the HADS and VTS scores of patients' relatives. Accordingly, a weak positive correlation was found between anxiety levels and tendency toward violence ($r=0.254$, $p<0.001$). Similarly, a weak positive correlation was found between depression levels and tendency toward violence ($r=0.218$; $p<0.001$) (Table 4).

As a result of the multiple linear regression analysis, the model was found to be statistically significant ($F=15.015$, $p<0.001$), with an explanatory power (R^2) of 0.075 This indicates that the anxiety and depression variables included in the model explain 7.5% of the total variance in the tendency toward violence. An analysis of the regression coefficients showed that anxiety scores were positively and significantly associated with a tendency toward violence ($\beta=0.193$, $p=0.001$). Similarly, depression scores were positively and significantly associated with a tendency toward violence ($\beta=0.121$, $p=0.036$). The model's intercept ($B=31.115$) indicates that even when anxiety and depression scores are zero, patients' relatives still exhibit some tendency toward violence (Table 5).

Participants' descriptive characteristics were compared with their HADS and VTS scores. Anxiety levels were significantly higher in females ($p<0.001$), single individuals ($p=0.030$), and those with a psychiatric history ($p<0.001$). Similarly, depression scores were significantly higher in participants with a psychiatric history ($p=0.001$). With respect to violent tendencies, no statistically significant differences were found among demographic groups ($p>0.05$) (Table 6). Additionally, the mean anxiety score of relatives who stated they were "not anxious at all" about their patient's condition was significantly lower than that of relatives who reported being "fairly anxious" or "very anxious" ($p=0.003$). Regarding tendencies to seek redress, mean anxiety and depression scores among relatives who stated "I will manage/do nothing" were significantly higher than among relatives who stated "I do not feel victimized" ($p<0.05$) (Table 6).

Table 3. Mean scores of participants on the Hospital Anxiety and Depression scale and Violence Tendency scale					
	Number of items	**Mean ± SD**	**Min**	**Max**	**Cronbach's Alpha**
Hospital Anxiety and Depression scale					
Anxiety	7	8.16±3.93	0	21	0.744
Depression	7	7.10±3.66	0	18	0.649
Violence Tendency scale					
Total	20	36.97±9.19	20	72	0.854
SD: Standard deviation					

Table 4. Relationship between participants' HADS scores and Violence Tendency scale scores		
		Violence Tendency scale
Anxiety	r	0.254
	p	**<0.001**
Depression	r	0.218
	p	**<0.001**
r: Pearson's correlation analysis, p: Significance level, p<0.001		

Table 5. Associations between Hospital Anxiety and Depression scale scores and participants' Violence Tendency					
	B	**SE**	**Beta**	**t**	**p**
Constant	31.115	1.166		26.674	<0.001
Anxiety	0.452	0.135	0.193	3.344	**0.001**
Depression	0.305	0.145	0.121	2.101	**0.036**
$R=0.275$. $R^2=0.075$. $F=15.015$. p<0.001 SE: Standard error, t: t-value, p: Significance level, p<0.001					

Table 6. Comparison of participants' descriptive characteristics with HADS and VTS scores					
Variable	**Category**	**Anxiety (mean ± SD)**	**Depression (mean ± SD)**	**VTS (mean ± SD)**	**p-value**
Gender	Female	8.65±3.85	7.04±3.74	36.41±9.00	A: <0.001 D: 0.643 V: 0.073
	Male	7.04±3.90	7.23±3.48	38.27±9.54	
Marital status	Married	7.67±4.01	6.90±3.60	35.98±8.75	A: 0.030 D: 0.407 V: 0.052
	Single	8.56±3.82	7.26±3.70	37.77±9.48	
Psychiatric history	Yes	10.32±3.90	8.89±3.70	37.26±7.12	A: <0.001 D: 0.001 V: 0.814
	No	7.92±3.86	6.89±3.60	36.94±9.41	
Level of concern	Not concerned at all	7.39±3.92	7.24±3.86	37.21±9.61	A: 0.003 D: 0.280 V: 0.170
	Slightly concerned	8.20±3.82	6.86±3.60	36.86±8.69	
	Moderately concerned	9.73±4.15	8.09±3.76	38.94±9.43	
	Very concerned	10.19±3.49	7.00±2.22	32.63±10.98	
Age category	18-25 years	8.51±3.78	7.07±3.66	38.40±9.81	A: 0.108 D: 0.051 V: 0.082
	26-35 years	7.46±4.12	6.28±3.69	36.76±8.03	
	36-45 years	8.12±4.16	7.40±3.76	35.86±9.49	
	46-55 years	8.76±3.35	8.32±3.08	34.56±8.38	
	56+ years	5.71±3.95	6.71±3.59	34.86±3.29	
Education level	Primary school	8.13±2.46	8.71±3.09	33.54±8.86	A: 0.336 D: 0.130 V: 0.310
	Secondary school	7.45±4.41	8.18± 4.38	38.91±10.49	
	High school	7.29±4.46	6.57±3.48	38.10±10.36	
	University	8.43±3.92	7.00±3.74	36.90±9.03	
	Postgraduate	8.22±3.65	7.18±3.42	37.11±7.94	
For multi-group variables (age, education), A: Anxiety, D: Depression, V: Violence Tendency p-values are presented, SD: Standard deviation					

Discussion

This study examined the levels of anxiety, depression, and violent tendencies among relatives of patients in the emergency department, and the associations between demographic factors and these variables.

Our study found that female participants and single individuals had significantly higher levels of anxiety. This finding is consistent with the existing literature. Indeed, in a study conducted by Yılmaz and Onan (6) on relatives of emergency department patients, women were reported to have higher levels of anger and anxiety compared to men. Similarly, Durna et al. (19) indicated that female employees are at a higher risk regarding stress and anxiety. When considered alongside the higher anxiety levels observed among single participants in our study, this trend suggests that the demographic characteristics of relatives of patients in the emergency department may play a determining role in their psychological well-being and, consequently, their propensity for violence.

The higher emotional reactivity in women, greater susceptibility to stressful life events, and gender-related differences in social support mechanisms are possible explanations for the higher anxiety levels observed in female participants (19). Among single individuals, factors such as lack of intra-family social support, feelings of loneliness, and reduced relational trust contribute to increased anxiety (13). Furthermore, in our study, participants with a psychiatric disorder (10.2%) had statistically significantly higher anxiety scores, suggesting that individuals with a psychiatric history may experience heightened anxiety levels when exposed to stressful healthcare environments. This result highlights the importance of closely monitoring patient relatives with a psychiatric history and providing them with psychosocial support, particularly in high-stress clinical areas such as emergency departments, to safeguard both the well-being of relatives and the safety of nurses.

In our study, the mean HADS-anxiety and HADS-depression scores were found to be at low-moderate levels (8.16±3.93; 7.10±3.66), while the mean VTS score was at a moderate level (36.97±9.19). Studies conducted in Türkiye indicate that relatives of patients in emergency departments often experience high levels of anxiety and stress. Yılmaz and Onan (6) reported that the intensity, uncertainty, and communication deficiencies in emergency departments are key factors increasing anxiety among patient relatives, while Aydemir et. al. (9) emphasized that long waiting times, difficulty accessing information, and inability to reach healthcare personnel elevate anxiety levels. Similar findings have been reported in the international literature. Hou et al. (7) noted that stress in the emergency department environment can increase both anxiety and violence tendencies, while Mol et al. (20) highlighted that uncertainty and insufficient information may lead to anger and increased violence learned helplessness among patient relatives. Accordingly, the dynamic, high-stress nature of emergency departments may increase anxiety and violent tendencies among relatives of patients if effective provision of information is not ensured. In this context, the critical roles of emergency department nurses and healthcare professionals in effective communication, crisis management, and empathetic approaches become evident. Regularly informing patients and their relatives, providing transparent explanations of processes, offering emotional support, and establishing a trustworthy communication environment are key components of nursing care and can directly contribute to reducing anxiety and tendencies toward violence. Therefore, strengthening healthcare professionals' communication skills, adopting care approaches that are sensitive to the psychological needs of patients' relatives, and increasing supportive nursing practices are crucial to reducing anxiety, stress, and tendencies toward violence in emergency departments.

The significant relationship found between "level of concern" and anxiety scores in the study ($p=0.003$) indicates that individuals experiencing high concern regarding their patient's condition are likely to have elevated anxiety. Literature also suggests that uncertainty in critical health situations increases anxiety among patient relatives (21). Moreover, individuals displaying passive attitudes toward seeking redress had higher anxiety and depression scores, which can be explained by loss of control and learned helplessness. Based on these findings, nurses in emergency departments can help reduce anxiety levels and prevent by establishing effective communication with patients' relatives, providing regular information, and offering emotional support. Strengthening nursing care in this way can reduce the risk of violence associated with anxiety, thereby supporting both the psychological well-being of relatives of patients and a safe working environment for healthcare professionals.

No significant differences were found in the descriptive characteristics of patient relatives with respect to their violence tendency scores ($p>0.05$). This finding is noteworthy when compared with other studies. While Yılmaz and Onan (6) reported that male individuals are more prone to physical violence, O'Brien et al. (12) reported that demographic characteristics alone are insufficient to explain violent behaviors and highlighted the significant influence of organizational, environmental, and psychosocial factors. The findings of our study suggest that violent tendencies may be more closely related to the conditions and quality of communication in the emergency department than to individual characteristics. Notably, patient relatives

identified "not being acknowledged" as the situation provoking the most anger, indicating that communication deficiencies are a critical trigger for violent behavior. Accordingly, the roles of emergency department nurses in communicating effectively, providing information, and involving relatives of patients in the care process are crucial in both reducing the risk of violence and ensuring a safe working environment.

In the study, a weak but statistically significant positive relationship was found between anxiety scores and violence tendency scores, and between depression scores and violence tendency scores (anxiety: $r=0.254$, $p<0.001$; depression: $r=0.218$, $p<0.001$). This finding is consistent with previous research indicating that elevated levels of anxiety and depression may impair emotional regulation, weaken anger control, and increase susceptibility to violence tendencies (5,22). High levels of anxiety may increase vulnerability to interpersonal conflicts, potentially contributing to the development of violence tendencies. Similarly, Roberton et al. (22) reported that both inadequate and maladaptive emotion regulation strategies may increase the likelihood of aggressive behaviors. Collectively, these findings suggest that heightened anxiety and depression are associated with violent tendencies, potentially by amplifying threat perception, increasing irritability, and lowering individuals' tolerance thresholds. Beyond individual psychological factors, the World Health Organization underscores that violence in healthcare settings is a complex, multidimensional issue shaped by organizational and systemic conditions, including excessive workload, prolonged waiting times, communication breakdowns, workplace stressors, and insufficient safety measures (23). Accordingly, effective strategies for violence prevention should extend beyond individual-level interventions and incorporate institutional and structural approaches. Within this framework, the results of the present study highlight the critical importance of the psychological well-being of patients' relatives in emergency departments, both to mitigate the risk of violence and to promote a safer working environment for nurses. Based on these findings, it is recommended that healthcare institutions implement regular psychosocial screening programs, expand stress management and anger regulation training, strengthen communication-based interventions between staff and patients' families, and enhance physical security measures in high-risk clinical settings. Given that emergency department nurses are the healthcare professionals who establish first contact with patients and their relatives, their role in identifying individuals experiencing psychological distress, developing appropriate communication strategies, and activating necessary psychosocial support mechanisms is key to preventing violence.

The finding in the multiple linear regression model that anxiety and depression were significantly associated with a tendency toward violence ($F=15.015$, $p<0.001$) indicates that psychological distress is related to the risk of violence in environments with high uncertainty and emotional burden, such as emergency departments. Although the explanatory power of the model was modest ($R^2=0.075$), it was statistically significant ($p<0.001$). This modest R^2 value is expected, given that the tendency toward violence in healthcare settings is a multifaceted phenomenon influenced by numerous systemic, environmental, and situational factors beyond individual psychological distress. Nevertheless, our findings indicate that levels of anxiety and depression remain significant independent factors associated with the psychological background of violent tendencies in the emergency department. Violence is not solely a result of individual psychological symptoms; it is a phenomenon shaped by a combination of factors including overcrowding, waiting times, lack of information, communication breakdowns, safety regulations, and organizational stressors. The World Health Organization also emphasizes that systemic and institutional measures are critical in preventing violence against healthcare workers (24).

The study's model coefficients show that anxiety has a stronger association with the tendency toward violence than depression does ($\beta=0.193$, $p=0.001$). Relatives of patients in emergency departments often experience feelings of loss of control, uncertainty, and perceived threat. Therefore, the resulting increases in anxiety levels are believed to reinforce individuals' perception of threat, increase irritability, and thus create an environment more prone to conflict and violent tendencies. Indeed, studies examining the prevalence and contextual triggers of violence in emergency departments reveal that such incidents are closely linked not only to "personal characteristics" but also to environmental conditions and interaction dynamics (6,25). The finding that depression also has an independent and significant positive association ($\beta=0.121$, $p=0.036$) suggests that the tendency toward violence should be evaluated not only for anxiety but also for factors such as negative mood, hopelessness, low coping capacity, and irritability. This was found to be consistent with current syntheses that address the psychosocial background of violence through the dimensions of emotion regulation and stress response (26). The high constant term ($B=31.115$) suggests that even when anxiety and depression are close to zero, the tendency toward violence persists at a baseline level. This situation indicates that violent behavior by relatives of patients is related not only to psychological variables but also to structural characteristics of the emergency department admission process (waiting times, overcrowding, lack of understanding of triage decisions), to security practices, and

to institutional communication factors. Policy documents and professional materials aimed at reducing violence need to provide a holistic framework that includes education, communication, environmental and safety regulations, reporting, and zero-tolerance approaches (27).

Study Limitations

This study has several limitations. First, it was conducted in a single secondary-level district state hospital, which may limit the generalizability of the findings to other healthcare settings. Second, the data were based on self-reported measures (HADS and VTS), which may be subject to social desirability bias and may reflect temporary psychological states. Future multicenter studies involving diverse socio-demographic populations are needed to enhance the generalizability and robustness of the findings.

Conclusion

This study identified a positive association between anxiety levels and violent tendencies among relatives of patients in emergency departments, highlighting the critical role of nursing practices in this context. The findings revealed that females, single individuals, and relatives with a psychiatric history experienced higher levels of anxiety, which could influence communication with nurses and potentially increase the risk of violence. Accordingly, it is recommended that emergency department nurses provide regular information, strengthen their empathetic and therapeutic communication skills, identify high-risk relatives early, and offer appropriate guidance and psychosocial support. Furthermore, to enhance nurse safety, it is important to strengthen intra-team communication, expand training in violence risk management, and establish safe working environments that support nurses. Future research should include multicenter studies that encompass diverse socio-demographic groups and healthcare settings, because the results could guide policy development for nursing services.

Ethics

Ethics Committee Approval: The research protocol was reviewed and approved by Karabük University Rectorate Social and Human Sciences Research Ethics Committee (decision number: 2025/01, date: 03.01.2025).

Informed Consent: Informed consent was obtained from all participants.

Footnotes

Authorship Contributions

Concept: S.Ö., Design: S.Ö., D.Y.G., Data Collection or Processing: S.Ö., Analysis or Interpretation: D.Y.G., Literature Search: S.Ö., Writing: S.Ö., D.Y.G.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Emergency Nurses Association. Emergency nursing: scope and standards of practice. 3rd ed. Burlington, MA: Jones & Bartlett Learning; 2023.
2. Bingöl S, İnce S. Factors influencing violence at emergency departments: patients' relatives' perspectives. Int Emerg Nurs. 2021;54:100942.
3. Lange S, Dąbrowska A, Koszucka K, Mędrzycka-Dąbrowska W. Resilience in family members of ICU patients: scoping review of the literature. J Clin Med. 2025;14:6794.
4. Spielberger CD. Manual for the state-trait anxiety inventory (STAI). Palo Alto (CA): Consulting Psychologists Press; 1983.
5. Neumann ID, Veenema AH, Beiderbeck DI. Aggression and anxiety: social context and neurobiological links. Front Behav Neurosci. 2010;4:928.
6. Yılmaz C, Onan N. Examination of anger expression and violence against healthcare professionals in the sample of emergency department patient relatives. ADYÜ Sağlık Bilimleri Derg. 2021;7:231-42. Turkish.
7. Hou Y, Corbally M, Timmins F. Violence against nurses by patients and visitors in the emergency department: an integrative review. J Adv Nurs. 2024;80:430-45.
8. Davey K, Ravishankar V, Mehta N, Ahluwalia T, Blanchard J, Smith J. A qualitative study of workplace violence among healthcare providers in emergency departments in India. Int J Emerg Med. 2020;13:33.
9. Aydemir I, Guclu R, Aydogan A. According to the emergency room staff the reasons for violence. J Ist Faculty Med 2020;83:60-8.
10. Katran M, Akdeniz YS, Ipekci A, Ikizceli I. A glimpse of violence against patients and relatives in the emergency department. Phnx Med J. 2020;2:160-7.
11. Spector PE, Zhou ZE, Che XX. Nurse exposure to physical and nonphysical violence, bullying, and sexual harassment: a quantitative review. Int J Nurs Stud. 2014;51:72-84.
12. O'Brien CJ, van Zundert AAJ, Barach PR. The growing burden of workplace violence against healthcare workers: trends in prevalence, risk factors, consequences, and prevention - a narrative review. EClinicalMedicine. 2024;72:102641.
13. Bandelow B, Michaelis S. Epidemiology of anxiety disorders in the 21st century. Dialogues Clin Neurosci. 2015;17:327-35.
14. Koyuncu A, Akman U, Yava A, Elagöz İ. Factors influencing the intention of family members of patients admitted to emergency departments to commit violence against health care professionals: a multicenter cross-sectional study. J Emerg Nurs. 2025;51:294-304.
15. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983;67:361-70.
16. Aydemir Ö, Kuey L, Küey L. Hastane anksiyete ve depresyon ölçeği (HADS) Türkçe uyarlaması: geçerlik ve güvenirlik çalışması. Türk Psikiyatri Derg. 1997;8:280-7.
17. T.C. Başbakanlık Aile Araştırma Kurumu. Aile içinde ve toplumsal alanda şiddet. Ankara: T.C. Başbakanlık Aile Araştırma Kurumu Yayınları; 1998.
18. Büyüköztürk Ş. Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum. Ankara: Pegem Akademi; 2021.
19. Durna BD, Çetin İ, Filiz M. A perception study on stress and stress management in healthcare professionals: Sivas sample hospital example. Van Sag Bil Derg. 2022;15:1-12. Turkish.

20. Mol MMC, Bakker EC, Nijkamp MD, Kompanje EJO, Bakker J, Verharen E. Relatives'perspectives on the quality of care in an intensive care unit: the theoretical concept of a new tool. Patient Educ Couns. 2014 ; 95:406-13.

21. Kang J, Cho YJ, Choi S. State anxiety, uncertainty in illness, and needs of family members of critically ill patients and their experiences with family-centered multidisciplinary rounds: a mixed model study. PLoS One. 2020;15:e0234296.

22. Roberton T, Daffern M, Bucks RS. Emotion regulation and aggression. Aggression and Violent Behavior. 2012;17:72-82.

23. Krug EG, Mercy JA, Dahlberg LL, Zwi AB. The world report on violence and health. Lancet. 2002;360:1083-8.

24. World Health Organization (WHO). Preventing violence against health workers. Avaible from: URL: https://www.who.int/activities/preventing-violence-against-health-workers

25. Chazel M, Alonso S, Price J, Kabani S, Demattei C, Fabbro-Peray P. Violence against nurses in the emergency department: an observational study. BMJ Open. 2023;13:e067354.

26. Emary EAM, Arifin SRM, Yusof MZ. A qualitative systematic review of healthcare practitioners' experience of workplace violence. Malays J Med Sci. 2024;31:51-61.

27. Wolf L, Perhats C, Delao A; 2019 ENA Position Statement Committee; Brim CB, Gentry JC, et al. Violence and its impact on the emergency nurse. J Emerg Nurs. 2020; 46:354-8.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Case Report
DOI: 10.4274/eajem.galenos.2026.24405
Eurasian J Emerg Med. 2026;25: 329-31

Erector Spinae Plane Block as the Primary Anesthetic Technique for Penetrating Chest Wall Injury: A Case Report

Sathyasuba Meenakshisundaram, Srinidhi Narayanan, Raghuraman M. Sethuraman

Sree Balaji Medical College & Hospital, Clinic of Anesthesiology, Chennai, India

Abstract

The erector spinae plane block (ESPB) is a simple regional anesthetic technique that is applied in various surgical procedures and has an excellent safety profile. Although primarily used for perioperative pain relief, it has also emerged as a useful opioid-sparing analgesic technique in emergency settings. Nevertheless, its application as the primary anesthetic technique for emergency surgical procedures is limited. We report a case of a penetrating lateral chest wall injury, which was successfully managed with an ultrasound-guided ESPB as the principal anesthetic technique for wound exploration and repair, obviating the need for general anesthesia or systemic opioids. The hemodynamics were stable overall, except for mild hypotension and tachycardia at initial presentation. This case highlights the potential use of ESPB as a safe alternative technique in emergency thoracic trauma surgery where neuraxial techniques or general anesthesia are contraindicated or unsafe.

Keywords: Erector spinae plane block, penetrating chest wall injury, emergency department

Introduction

Penetrating thoracic injuries present unique challenges in overall management; the anesthetic management poses additional challenges for the surgeries in these patients, as they require preserving respiratory function in addition to avoiding anesthetic agents/drugs that may impair ventilation (1). The erector spinae plane block (ESPB) has gained widespread popularity since its introduction a decade ago and has been applied in various surgical procedures. It achieves a reliable spread to the dorsal rami, while spread to the ventral rami is variable. Nevertheless, it has been shown to be an effective technique for pain relief in various surgeries, including breast, thoracic, and abdominal procedures.

ESPB is now increasingly employed in emergency department (ED) settings, especially as an opioid-sparing analgesic technique in various conditions (2). Furthermore, the continuous ESPB technique has demonstrated sustained analgesia in extensive thoracic trauma surgeries (3,4). A recent study observed that the landmark ESPB administered at a prehospital setting was effective for pain relief in patients with blunt chest trauma (5). Although ESPB has been applied for surgical anesthesia in a few elective surgical procedures (6,7), it is very limited in emergency surgeries. Balaban et al. (7) reported a case of subcostal stab wound repair under ESPB in addition to two elective procedures, in their case series. We report a case of penetrating chest wall injury wherein the ESPB was used as a sole technique for hemostasis and chest wall repair.

Case Report

A 19-year-old male presented to the ED with a penetrating injury to the right posterolateral chest wall following a fall from a height onto a sharp metal rod. On arrival, he was conscious, agitated, and tachypneic (respiratory rate 32/min), with a blood pressure of 94/60 mmHg, a heart rate of 110 beats per minute, and severe pain, with a visual analog scale (VAS) score of 9/10. Active bleeding was partially controlled by external compression. The wound measured 18X14X2 cm in length, breadth, and depth. The wound extended only to the chest wall musculature, with no evidence of injury to the heart, pericardium, lungs, pleura, or other mediastinal structures. Intravenous fluids (1 L) were administered

Corresponding Author: Raghuraman M. Sethuraman Prof, Sree Balaji Medical College & Hospital, Clinic of Anesthesiology, Chennai, India
E-mail: drraghuram70@gmail.com **ORCID ID:** orcid.org/0000-0001-8464-7458

Cite this article as: Meenakshisundaram S, Narayanan S, Sethuraman RM. Erector spinae plane block as the primary anesthetic technique for penetrating chest wall injury-a case report. Eurasian J Emerg Med. 2026;25: 329-31.

Received: 22.11.2025
Accepted: 01.06.2026
Published: 11.06.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

for hemodynamic stabilization. The e-FAST performed by the emergency physician on the patient's arrival was unremarkable. Chest X-ray and bedside ultrasonography, performed by the specialist in the radiology department, ruled out pneumothorax or hemothorax. The wound involved intercostal muscles without pleural breach and required urgent exploration and repair in the emergency operating room, as the patient had active bleeding and intractable pain (Figure 1).

After obtaining informed consent (for using the clinical image for research purposes), with patient in the left lateral decubitus position, under conscious sedation with intravenous midazolam 1.5 mg, a single-shot ultrasound-guided ESPB was performed with high frequency linear probe (11 to 13 MHz) at the T5 level using an in-plane approach, by placing the 22 G 3.5-inch needle, between the erector spinae muscle and the TP of vertebrae ensuring longitudinal anesthetic spread. After negative aspiration was confirmed, 20 mL of 0.5% ropivacaine was injected into the plane between the erector spinae muscle and the transverse process. A sensory block (confirmed by pin-prick) from T3 to T8 dermatomes developed within 15 minutes. The procedure was carried out in the block room with continuous monitoring of vital signs and appropriate resuscitation equipment in place. The entire surgery proceeded uneventfully under ESPB and mild sedation, without additional sedation or systemic opioids. The contingency plan was to use general anesthesia with an endotracheal tube and controlled ventilation in case of block failure. The wound was closed after confirming hemostasis (Figure 2). Intraoperative hemodynamics remained stable, and the patient reported minimal pain (VAS 2/10) postoperatively. He recovered uneventfully and was discharged after 48 hours.

Discussion

This case illustrates the expanding role of the ESPB as a valuable anesthetic and analgesic tool in thoracic trauma, as it effectively manages chest wall pain while preserving respiratory function.

The application of the ESPB has increased markedly over the past few years, as it is easy to perform and effective at reducing analgesic requirements, with minimal theoretical complications. A review article observed that ESPB was applied for analgesia in various acute painful conditions presenting in the ED setting, such as acute pancreatitis, burn injuries, herpes zoster, renal colic, and mechanical pain, with rib and spine fractures being the most common (2).

Recently, two case reports demonstrated successful continuous ESPB for sustained perioperative analgesia in extensive thoracic trauma surgeries (3,4). A recent study concluded that ESPB resulted in reductions in pain scores and opioid consumption,



Figure 1. Penetrating chest wall injury with bleeding



Figure 2. Sutured wound after hemostasis

without adverse effects on hemodynamics, in patients with blunt chest trauma. Notably, the ESPB was applied in a prehospital setting by the landmark technique without ultrasound guidance (5). A few case reports applied ESPB as a sole anesthetic technique in gastrostomy (6), and in minor breast surgeries (7). Balaban et al. (7) used ESPB in a case of subcostal stab wound repair in their case series (7). Thus, we believe that available reports on emergency surgical procedures are very limited.

Compared with neuraxial and paravertebral techniques, the ESPB offers several advantages, including ease of administration, superficial needle placement, and reduced risk of complications such as hypotension, dural puncture, or pneumothorax. Compared with local infiltration, ESPB provides wider, more consistent, and longer-lasting dermatomal coverage. Administration of general anesthesia may be unsafe in some cases because of comorbidities or a full stomach.

In this patient, the ESPB provided both surgical anesthesia and postoperative analgesia without hemodynamic instability or procedural difficulties. While a single-shot block is suitable for minor procedures, a continuous catheter may be preferred for prolonged or extensive chest wall injuries to ensure sustained pain relief (3,4).

Although anesthesiologists were involved in this case, as per our institutional protocol in the emergency operating room, this case report demonstrates that emergency/trauma physicians can also apply this technique, owing to its simplicity and safety in such settings. Furthermore, ESPB is a safe and simple regional anesthetic technique that may reduce opioid use.

Conclusion

A single-shot ESPB may be considered in selected cases as an effective anesthetic technique for chest wall surgical procedures in the emergency setting. It is safe, easy to perform, and offers reliable anesthesia and analgesia with minimal physiological disturbance. However, continuous ESPB, preferably administered with general anesthesia, is required for extensive or major procedures.

Ethic

Informed Consent: The informed consent form was obtained from the patient for this case report.

Footnotes

Author Contributions

Surgical and Medical Practices: S.M., S.N., R.M.S., Concept: S.N., R.M.S., Design: S.M., S.N., R.M.S., Data Collection or Processing: S.N., R.M.S., Analysis or Interpretation: S.M., R.M.S., Literature Search: S.M., R.M.S., Writing: S.M., R.M.S.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Saillant NN, Sein V. Management of severe chest wall trauma. J Emerg Crit Care Med. 2018;2:41.
2. Abdelhamid K, ElHawary H, Turner JP. The use of the erector spinae plane block to decrease pain and opioid consumption in the emergency department: a literature review. J Emerg Med. 2020;58:603-9.
3. Dmytriiev D, Dîrzu DS, Melnychenko M, Eichholz R. Erector spinae plane block for affective and safe analgesia in a patient with severe penetrating chest trauma caused by an explosion in the battlefield. Clin Case Rep. 2022;10:e6433.
4. Sen S, Hernandez N, Langridge XT, Lamaris GA, Ge MA, De Haan JB. Continuous erector spinae plane block for acute pain control in chest wall reconstruction of extensive traumatic degloving injury. Trauma Case Rep. 2021;32:100415.
5. Basyouny KAR, Abd El-Kader M, Mahmoud MF, Abdelhamid BM, Waheeb MME. Erector spinae plane block for pain management in blunt chest trauma in military prehospital medicine, an interventional study. Egypt J Anaesth. 2024;40:351-7.
6. Bagaphou CT, Piraccini E, Norgiolini L, Ciabucchi C, Carsena V, Evoli LP, et al. Anesthesia experience for open gastrostomy with ultrasound-guided erector spinae plane block: a case report. Case Rep Anesthesiol. 202;2020:5413848.
7. Balaban O, Aydin T, Yaman M. Is ultrasound guided erector spinae plane block sufficient for surgical anesthesia in minor surgery at thoracal region? J Clin Anesth. 2018;47:7-8.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
DOI: 10.4274/eajem.galenos.2026.90698
Review
Eurasian J Emerg Med. 2026;25: 332-41

Advances in Acute Management and Evolving Clinical Practices in Aluminum Phosphide Poisoning: Insights from Recent Case Studies

Saeed-Aghebat Bekheir[1], Amir Baghaei[2], Hamed Haghi-Aminjan[1], Maryam Baeeri[1], Shirin Hekmatirad[1], Anita Jahanpanah[1], Omid Mehrpour[3], Mohammad Abdollahi[1]

[1]Tehran University of Medical Sciences Faculty of Pharmacy and Pharmaceutical Sciences Research Center (PSRC), Department of Toxicology and Pharmacology, Tehran, Iran
[2]Alborz University of Medical Sciences Faculty of Pharmacy, Department of Toxicology and Pharmacology, Karaj, Iran
[3]Southern Methodist University, Data Science Institute, Dallas, Texas, USA; Scientific Unlimited Horizon, Tucson, USA

Abstract

Aluminum phosphide (AlP) is a pesticide widely used by farmers in certain Asian countries for its high efficacy in protecting crops from pests. However, misuse in both intentional and unintentional poisoning cases has raised major public health concerns. Despite extensive research into treatments for AlP poisoning (AlPP), no specific antidote exists, and management remains largely supportive. This review examined successful treatment cases reported in the literature. A targeted search was conducted using the keywords "AlP", "phosphine", "antidote", "case report", and "treatment" in the scientific databases Scopus, Web of Science, and PubMed. The identified treatments include disinfection of the digestive tract (including gastric disinfection); prevention of phosphine gas release; administration of coconut oil and olive oil; magnesium sulfate; thyroid hormones; extracorporeal membrane oxygenation; continuous renal replacement therapy (including continuous veno-venous hemofiltration); high-dose insulin; red blood cell exchange; dihydroxyacetone; and antioxidant therapy. These therapeutic approaches may serve as supportive therapy by reducing the harmful effects of AlP. This article evaluates medications and chemical substances proposed for the targeted treatment of AlPP and examines their protective mechanisms and the primary consequences of their use.

Keywords: Aluminum phosphide, rice tablet, phosphine, antidote, case report, poisoning

Introduction

Aluminum phosphide (AlP) is among the most common and cost-effective pesticides and is widely used in the grain-preservation industry. It is often encountered as rice tablets containing about 56% AlP (1). Once in contact with moisture, water, or gastric contents, AlP generates phosphine gas (PH_3), a highly poisonous, hazardous entity (2,3). It is a significant public health problem: thousands of people worldwide die every year from pesticide poisoning. AlP poisoning (AlPP) has become one of the most common methods of suicide involving pesticides in several countries. In specialty clinics treating intoxications, AlPP accounted for approximately 76% of cases per year, with a mortality rate of two-thirds among affected individuals (4). This is especially dangerous in developing countries like India, Sri Lanka, and Iran, where the rate of suicide by AlP is increasing (5-7).

As shown in the Figure 1 AlP toxicity results from the release of phosphine gas (3). This gas blocks oxidative phosphorylation by inhibiting mitochondrial cytochrome oxidase, ultimately causing cellular hypoxia and circulatory collapse (8). PH_3 inhibits cellular metabolism, which manifests in symptoms like nausea, diarrhea, colic, hypotension, and life-threatening cardiovascular disorders such as congestive heart failure and arrhythmias (Figure 2).

Corresponding Author: Prof. Mohammad Abdollahi, Tehran University of Medical Sciences Faculty of Pharmacy and Pharmaceutical Sciences Research Center (PSRC), Department of Toxicology and Pharmacology, Tehran, Iran
E-mail: Mohammad@TUMS.Ac.Ir **ORCID ID:** orcid.org/0000-0003-0123-1209

Cite this article as: Bekheir S-A, Baghaei A, Haghi-Aminjan H, Baeeri M, Hekmatirad S, Jahanpanah A, et al. Advances in acute management and evolving clinical practices in aluminum phosphide poisoning: insights from recent case studies. Eurasian J Emerg Med. 2026;25: 332-41.

Received: 22.11.2025
Accepted: 02.06.2026
Published: 26.06.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

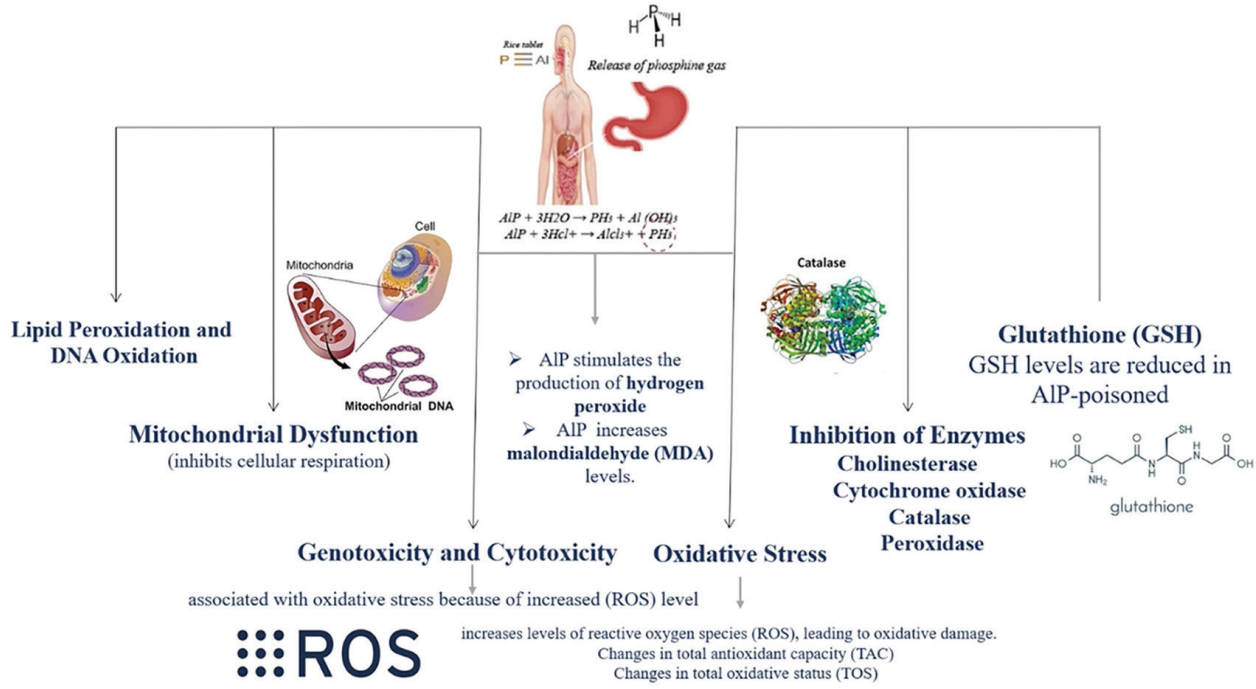


Figure 1. The mechanisms of acute aluminum phosphide poisoning

AIP: Aluminum phosphide, ROS: Reactive oxygen species

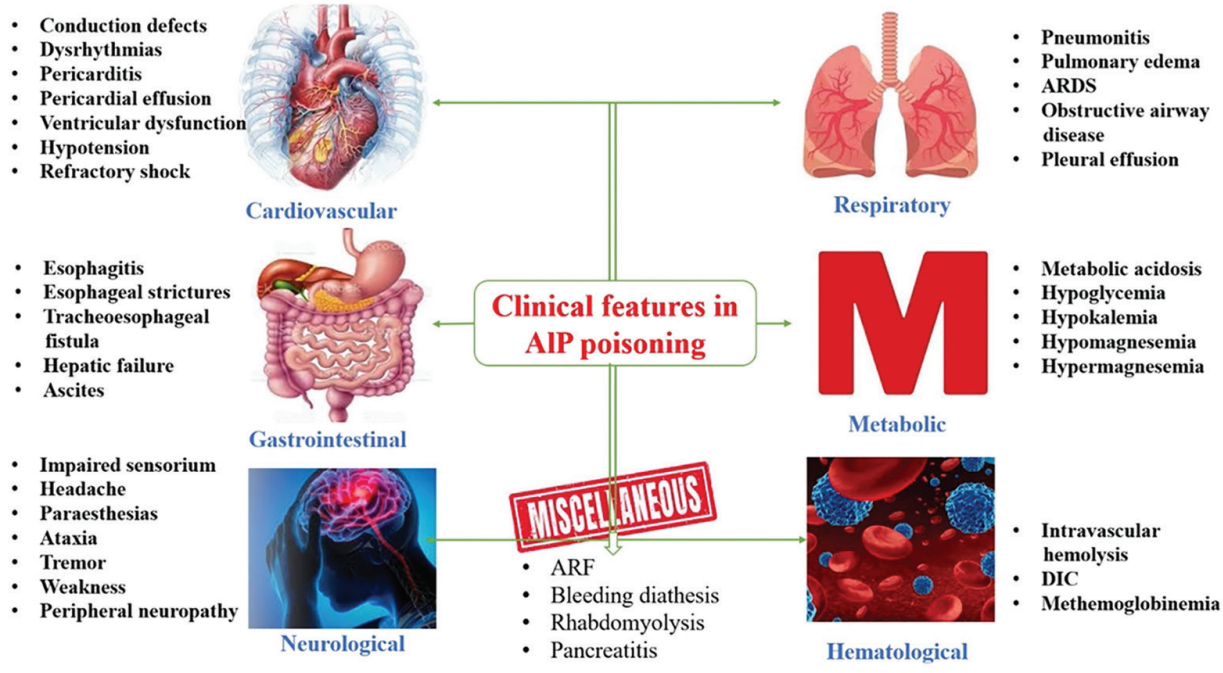


Figure 2. Clinical symptoms of acute aluminum phosphide poisoning

AIP: Aluminum phosphide, ARDS: Acute respiratory distress syndrome, ARF: Acute respiratory failure, DIC: Disseminated intravascular coagulation

Evolving respiratory symptoms can range from dyspnea to acute respiratory distress syndrome (ARDS) and pulmonary edema (9).

AlPP patients may present with manifestations ranging from mild gastrointestinal upset to severe cardiovascular and respiratory failure. Intravascular hemolysis, however, has also been recognized as a rare complication of this poisoning (9). Currently, there is no specific antidote for AlPP; treatment remains largely supportive. Management focuses on correcting metabolic acidosis and addressing severe symptoms resulting from poisoning (Figure 3). Despite the high mortality rates associated with severe cases, ranging from 60% to 85%, there have been instances of successful recovery through prompt medical intervention (10-12).

Accordingly, in this article we review and summarize recent guidelines and findings on the management of acute AlPP.

Materials and Methods

A comprehensive search of the PubMed, Scopus, and Web of Science databases was conducted to identify relevant studies. The search strategy included the keywords: "AlP", "phosphine", "antidote", "case report", and "treatment". Case reports and articles describing successful treatment approaches in human studies for AlPP were reviewed. After the initial retrieval, duplicate records were removed, and the remaining articles were screened for relevance to the study objective. No time limit was applied to the articles, and all case reports describing successful treatments of AlPP were reviewed. Excluded were non-clinical or non-human studies (such as reviews, animal studies, and in vitro research); reports involving phosphide compounds other than AlP; articles with unclear outcomes or fatal cases; and reports lacking sufficient details on the treatment approach or clinical course.

Results

Gastric Decontamination

Since the release of PH_3 begins when AlP interacts with water and stomach acid, researchers have studied various methods to reduce this release.

Potassium Permanganate

There are contradictory reports, with some asserting that gastric washout using potassium permanganate can convert PH_3 into a harmless phosphate through oxidation (13). Furthermore, Sanaei-Zadeh (14) noted that potassium permanganate is a potent oxidizing agent and documented instances of hemolysis and methemoglobinemia following AlPP. Initially, these patients were managed by administering potassium permanganate through gastrointestinal lavage (14,15). Sanaei-Zadeh and Marashi (16)

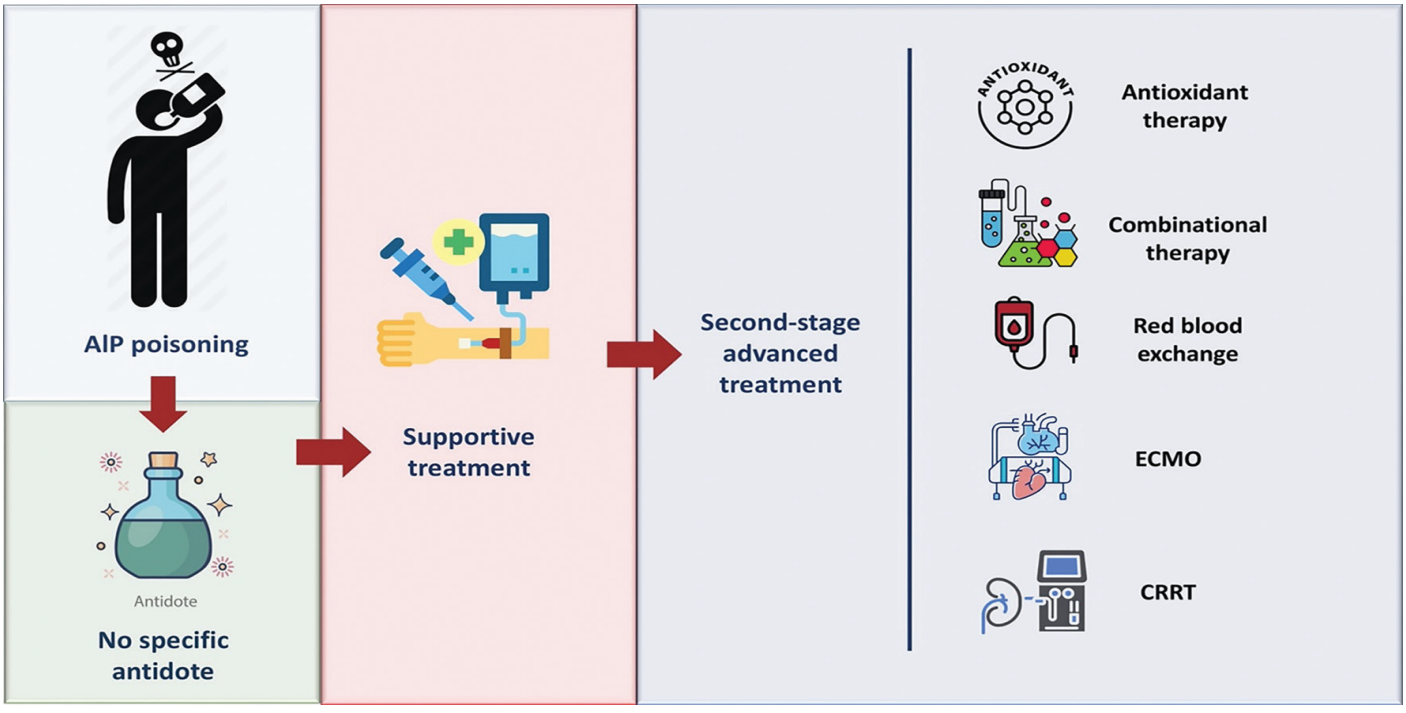


Figure 3. Aluminum phosphide (AlP) poisoning has no specific antidote; initial management is supportive, with escalation to second-stage advanced therapies in nonresponsive cases

CRRT: Continuous renal replacement therapy, ECMO: Extracorporeal membrane oxygenation

state that potassium permanganate is a chemical that easily dissolves in water and can prevent the release of PH_3 gas.

Liquid Paraffin and Castor Oil

Laboratory studies have confirmed that liquid paraffin and vegetable oils can prevent the further release of PH_3 gas (17). Therefore, the use of vegetable oils or liquid paraffin alone for gastric decontamination after acute AlPP is recommended, although gastric lavage with vegetable oils is technically possible. Sanaizadeh and Marashi propose that the management of castor oil effectively prevents the continued release of PH_3 upon interaction with the moist digestive system, accelerates the movement of the digestive system, and facilitates gastrointestinal elimination of the toxic agent (16).

Olive Oil

According to a study (18) gastric lavage with potassium permanganate 1:1000, activated charcoal, and coconut oil can be used as the first step in an emergency to help reduce the rate of PH_3. In another study, it has been suggested that gastric lavage with sodium bicarbonate and coconut oil may represent a feasible supportive approach (19). Another study announced that zinc phosphide poisoning was effectively treated by performing gastric washout using olive oil and providing other supportive therapies (20). Olive oil has been proposed as a potential adjunctive agent that may reduce PH_3 release and improve clinical outcomes.

Coconut Oil

Coconut oil contains a high concentration of saturated fatty acids. It has been reported that coconut oil can prevent the release of PH_3 from AlP (17). In the case of AlPP, a 28-year-old man was administered coconut oil six hours after consuming 12 grams of AlP, showing that coconut oil reduces the expected amount of AlP absorption (21). This study proposed coconut oil as a potentially life-saving and beneficial supplementary treatment for AlPP.

Another investigation involved the treatment of thirty-three patients who had been poisoned with AlP. These patients had thorough gastric lavage, including sodium bicarbonate and coconut oil (22). Based on these findings, the authors reported that coconut oil and other therapeutic interventions may improve survival outcomes.

A dissimilar study on seven patients seriously affected by AlPP found that they were managed with supportive interventions, including permanganate, coconut oil, sodium bicarbonate, and gastric washout with diluted potassium (19). Based on a viability rate of 51.14% among poisoned patients, the authors propose that coconut oil could be recognized as a viable treatment option in conjunction with other therapeutic interventions for managing cases of poisoning.

Magnesium Sulfate

Several studies investigating the effect of magnesium sulfate on AlPP have reported contradictory results regarding its therapeutic efficacy. According to some studies, AlPP does not decrease magnesium levels in these patients, so magnesium sulfate does not improve the patient's condition (22,23). In contrast, other studies exhibited that magnesium levels decrease in AlPP, thus confirming the hypothesis that enhancing magnesium levels can be considered an effective treatment for AlPP (24). In a study of antioxidant effects, magnesium was evaluated in several patients who were poisoned by AlP (25). After demonstrating that oxidative stress induced by AlP during the initial stage of intoxication was responsible for increased lipid peroxidation and decreased glutathione (GSH) levels, the researchers reported meaningful amelioration among patients receiving magnesium. Therefore, magnesium can be considered an adjunctive treatment for reducing oxidative stress caused by AlPP.

Liothyronine

Thyroid hormones exert short- and long-term effects on mitochondria. The administration of thyroid hormones in the management and therapy of AlPP is primarily attributed to the mitochondrial-safeguarding properties of these substances (26). Following the initial interventions, a dosage of 50 µg of liothyronine was delivered to twelve patients who had been poisoned with AlP (27). Oral administration of liothyronine significantly improved systolic blood pressure, arterial pH, and total thiol molecule concentration. Furthermore, liothyronine reduces lipid peroxidation, increases catalase activity, and preserves total antioxidant capacity. Therefore, liothyronine may be a promising adjunctive therapy for AlP-poisoned patients.

Digoxin

Cardiogenic shock is a major cause of death in AlPP. Digoxin can have a protective role in such cases. For example, an 18-year-old female with acute AlPP and severe cardiac failure was successfully treated with digoxin. The patient received an initial digoxin dose (0.5 mg), followed by 0.5 mg every 6 hours on the first day and 0.25 mg daily thereafter. She fully recovered and was discharged 10 days later. This case suggests that digoxin may be beneficial in managing cardiogenic shock in AlPP (28). Another successful case was a 21-year-old male who consumed a 3-gram tablet of AlP and was successfully treated by administering a combination of antioxidants and digoxin (29).

Intra-aortic Balloon Pump

The intra-aortic balloon pump (IABP) is among the earliest and most prevalent forms of mechanical circulatory support. The device operates through external counterpulsation, employing systolic unloading and diastolic augmentation of aortic pressure to enhance hemodynamics. Despite providing inferior hemodynamic support compared to contemporary mechanical circulatory support devices, IABP may still be the preferred option in suitable circumstances because of its relative ease of insertion and removal, minimal vascular access requirements, and a superior safety profile. IABPs improve myocardial function by optimizing oxygen delivery and reducing oxygen consumption (30). Recently, this method has been used to treat and manage cardiogenic shock due to AlPP alone or in combination with other therapeutic approaches, and promising therapeutic outcomes have been achieved (31).

Three patients with AlPP developed refractory shock despite receiving IV fluids and vasopressors. IABP was attempted, but only one survived (32). Mehrpour et al. (33) reported two cases, a 17-year-old man and a 21-year-old woman, successfully treated with IABP for cardiogenic shock induced by AlPP. Mehrpour et al. (34) also reported a successful case of AlPP treated with IABP. A 24-year-old woman presented with refractory AlP-induced cardiogenic shock and ARDS after ingesting 3 g of AlP. She received gastric lavage, sodium bicarbonate, and intensive care, but her LVEF dropped to <20%. An IABP was inserted, stabilizing her vital signs. By day 20, her LVEF improved to 50%, demonstrating that IABP may be a promising intervention for severe AlPP. Moreover, Siddaiah et al. (35) reported another successful treatment of AlPP using IABP.

Extracorporeal Membrane Oxygenation

Extracorporeal membrane oxygenation (ECMO) is a type of circulatory support used in patients with heart failure, respiratory failure, or both who have failed standard therapies. It provides the body with sufficient time to detoxify and improve organ function (36). A study conducted in India revealed that doctors' use of ECMO resulted in a significant decrease in mortality among poisoned patients. The mortality rate dropped from 88% in untreated cases to 33% when ECMO was employed, highlighting the remarkable effectiveness of this treatment (37).

In another study conducted in Nepal, the patient's hemodynamic status and general condition gradually improved within two days. The patient was transferred to the intensive care unit on the sixth day of ECMO and was discharged from the hospital on the tenth day after ECMO (38).

The authors Mohan et al. (37) The initial study documented the effective treatment of AlPP with refractory cardiogenic shock using ECMO. Hassanian-Moghaddam et al. (39) conducted a study demonstrating a case of AlPP, which was completely cured after four days of treatment with ECMO. Further observational evidence has emphasized the importance of initiating ECMO promptly before profound deterioration in ventricular function. Mohan et al. (40) reported that early veno-arterial ECMO support was associated with improved survival and reversal of myocardial dysfunction in patients with severe AlPP. All of the above studies show that ECMO is a suitable intervention for patients with poor hemodynamic status who have not responded to other conventional treatments.

Continuous Renal Replacement Therapy

Continuous renal replacement therapy (CRRT) is frequently employed to provide renal support for critically sick patients suffering from acute kidney injury, particularly those who are experiencing hemodynamic instability (41). In one study, CRRT was started within two hours of presentation in two poisoned patients after initial reports indicated elevated lactate. CRRT was performed using bicarbonate-based fluids and continued until metabolic acidosis resolved and serum lactate levels normalized; both patients survived. It is thought that CRRT can correct severe metabolic acidosis and stabilize hemodynamic status, helping improve the condition of poisoned patients (42). CRRT is superior to intermittent hemodialysis in these patients and can be used for long-term, hemodynamically unstable patients for whom conventional hemodialysis is relatively contraindicated. In addition, CRRT helped maintain the metabolic environment and resolve the shock state until AlP excretion (43). It is recommended that such extracorporeal treatments be used for patients at high risk of mortality, together with supportive therapies until the hemodynamic condition improves.

Peritoneal Dialysis

Bashardoust et al. (44) reported two severe AlPP cases successfully treated with peritoneal dialysis, highlighting its potential role in correcting metabolic acidosis and improving survival outcomes.

Continuous Veno-venous Hemofiltration

This method treats acid-base and electrolyte disorders in patients with unstable hemodynamics. Several reports describe the successful use of hemofiltration in cases of drug poisoning complicated by hypotension and metabolic acidosis (45). A case report describes a patient who underwent continuous veno-venous hemofiltration for AlPP. Unfortunately, this patient died due to ventricular tachycardia (46). Nasa et al. (42) reported two patients with acute albinism who survived after CRRT. In a separate investigation, hemofiltration was initiated immediately for refractory hypotension and metabolic acidosis; a serum creatinine level of 1.1 mg/dL was measured, so the patient underwent continuous veno-venous hemofiltration for 32 hours. The patient was discharged healthy on the fourth day of hospitalization (47). Notably, hemofiltration should be started before multiorgan

failure occurs (MOF) in our patient, because prolonged periods of metabolic acidosis and hypotension resulting from irreversible cardiac complications and progression of MOF cause high mortality.

High-dose Insulin

In the previous experiences of AlPP, almost all patients poisoned with AlP died of severe hypotension and metabolic acidosis; in this case report study, this patient was successfully managed using glucose-insulin-potassium (GIK) (48). Sedaghattalab (48) reported the successful treatment of a 30-year-old Iranian woman with severe AlPP, presenting with hypotension, shock, and severe metabolic acidosis, using high-dose regular insulin and hypertonic dextrose.

In a study conducted on 88 patients poisoned with AlP, it was shown that 44 of them received insulin infusion, which supports our experience (49). Another experimental study of 60 patients with rice tablet poisoning showed that the durability rate was significantly higher in the GIK group (50). GIK therapy may represent a promising adjunctive strategy in acute AlPP; however, further clinical studies are required before routine incorporation into treatment protocols.

Red Blood Cell Exchange

An animal study showed that infusion of fresh RBCs ameliorated metabolic acidosis and increased survival in mice exposed to AIP. Rahimi et al. (51) stated that hemoglobin has a powerful buffer capacity, enabling red blood cells to show the capacity to preserve blood pH. On the other hand, Zamani et al. (52) stated that an adult patient with AlPP did not respond to traditional treatment and recovered after a complete blood exchange procedure. In addition, some case reports have shown that G6PD deficiency can play a protective role in AlPP, as G6PD insufficiency leads to the breakdown of red blood cells under oxidative stress (52-54). Therefore, it was suggested that removing damaged red blood cells may be helpful in patients exposed to AIP and may be indicated as a potential adjunctive therapeutic option for AlPP.

Hyperbaric Oxygen

Hyperbaric oxygen therapy (HBOT) involves exposure to pure oxygen at elevated atmospheric pressure (55). Several animal studies have investigated the effectiveness of HBOT in rats. The results have shown that this method can be an effective treatment for mild poisoning. This approach may serve as an effective therapeutic option for minor poisonings. This approach may enhance the survival duration of mice exposed to AlPP, although it may not decrease the fatality rate (56). The first reported case of successful treatment of poisoning with HBOT involved a patient who had suffered accidental AlPP. Initially misdiagnosed as carbon monoxide poisoning, the patient received comprehensive treatment, including HBOT, and fully recovered, marking the first reported clinical case of HBOT use in AlP intoxication (57).

Dihydroxyacetone

Dihydroxyacetone is synthesized as dihydroxyacetone phosphate during glycolysis within the cell, and it plays a crucial role in the production of adenosine triphosphate (ATP). PH_3 toxicity primarily occurs through inhibition of ATP generation in cells. However, the entry of DHA into the glycolysis pathway can avert an energy crisis (ATP depletion) in cells and prevent cell death. This is crucial for organs particularly vulnerable to ATP depletion, such as the heart and brain. A recent animal study demonstrated that DHA substantially reduced mortality (58).

A case report showed that in addition to the usual management of AlPP, patients received DHA. Administering DHA via gavage at a dose of 7 g in 50 ml of sodium bicarbonate, twice at 1-hour intervals, improved clinical symptoms. This is the first case report to emphasize the effective treatment of AlPP with DHA (59). However, it was found that administration of DHA in poisoned patients significantly improved severe acidosis and AlP-induced PaO_2 (60).

Antioxidant Therapy

After exposure to AlP, a decrease in GSH plasma concentration has been observed and reported (25). N-acetylcysteine (NAC) can increase GSH reserves. Regarding the efficacy of NAC in AlPP, there are conflicting reports, and only one documented case has been reported in which NAC delivery, as part of treatment, failed to save the patient (61).

Nevertheless, Azad et al. (62) reported that NAC injection can substantially increase survival in a mouse model. Consequently, the administration of NAC can be considered a therapeutic agent in addition to a multi-therapeutic strategy.

Marashi et al. (63) recommended that the use of coenzyme Q10 (CoQ10) as an antioxidant could be regarded as an alternative therapy approach. According to their assertion, relying on prior research conducted on patients with heart failure with a similar condition to AlPP, CoQ10 can also increase the systolic function of the heart and help save the poisoned person (64-66).

Recently, vitamin E has been recognized as an effective treatment for managing people with AlPP. A 21-year-old male who consumed a 3-gram tablet of AlP was successfully treated with a combination of antioxidants. This treatment included a slow intravenous infusion of 1000 mg of vitamin C every 12 hours, an intramuscular injection of 400 units of vitamin E, and an oral NAC. The NAC dosage was 70 mg/kg every 4 hours after an initial loading dose of 140 mg/kg for a maximum of 17 doses (29).

When combined with additional supportive treatments, vitamin E, 400 mg BD intramuscularly, dramatically reduced mortality in subjects exposed to AlP (67).

Table 1 summarizes the key therapeutic interventions and clinical outcomes from the reviewed case studies.

Table 1. Therapeutic interventions and clinical outcomes in cases of aluminum phosphide poisoning: a review of case studies

Treatment/intervention	Study model, year, country	Gender, age, and amount of AlP	Outcome, reference
CRRT	Case report 2013 India	Female 30 half tablet (1.5 gm) Man 24 one tablet	Early initiation of CRRT, in addition to other supportive treatments, resulted in the survival of both poisoned patients (42).
Continuous veno-venous hemofiltration	Case report 2016 Türkiye	Female 24 two tablets	The patient's hemodynamic condition improved during hemofiltration, which increased blood pressure and gas levels for up to 32 hours. She was discharged from the hospital on the fourth day of her stay (47).
ECMO	Case report 2018 Nepal	Female 67 two tablets	The patient who had been poisoned was released from the intensive care unit on the sixth day following ECMO. She was discharged from the hospital on the tenth day after ECMO (38).
IABP	case report 2019 Iran	Two cases of poisoning	This case study confirmed the successful treatment of cardiogenic shock due to AlP poisoning in a 17-year-old man and a 21-year-old woman using the IABP procedure (33).
DHA (7 gr in 50 mL sodium bicarbonate, gavage)	Case report 2020 Iran	Two cases of poisoning	The clinical symptoms were relieved, and the patients were discharged after administering DHA (7 g in 50 mL of sodium bicarbonate via gavage) (59).
Magnesium sulfate, trimetazidine IABP, and ECMO	Case report 2021 Nepal	Female 17 two tablets	She was shifted to the general division on the 11th day of her admission (68).
ECMO and CRRT	Case report 2021 China	Female 15 one tablet (3.2 g)	The poisoned patient was successfully treated with ECMO along with CRRT (69).
Sodium bicarbonate (50 ccs), aloe vera syrup (1-2 L), and olive oil (150 cc)	Case report 2021 Iran	Man 33 three tablets	Combination therapy was a successful experience in treating acute AlPP (70).
Red blood cell exchange	Case report 2021 Türkiye	Female 15 500 mg	Despite the lack of improvement in the patient's hemodynamic condition with supportive treatment, she achieved successful treatment by automated red blood cell exchange (71).
Olive oil	Case report 2021 Iran	Man 33 three tablets	Research from this study suggests that using olive oil as a gastric lavage may be beneficial (72).
High-dose insulin	Case report 2022 Iran	Female 33 unclear	Treatment with these drugs was successful, and this patient was saved with glucose-insulin-potassium (48).
VA-ECMO and IABP	Case report 2023 Russia	Man 20 unclear	The patient was discharged after 34 days of treatment (31).
HBOT	Case report 2024 Czech Republic	Male 40 accidental	The patient was treated entirely with HBOT and discharged (57).

CRRT: Continuous renal replacement therapy, ECMO: Extracorporeal membrane oxygenation, IABP: Intra-aortic balloon pump, VA-ECMO: Veno-arterial membrane oxygenation, HBOT: Hyperbaric oxygen therapy

Conclusion

AlPP poses a significant global health challenge because of its high mortality rate and the lack of a definitive antidote. Managing it is complex because exposure to moisture rapidly releases phosphine gas, which can cause severe systemic toxicity.

Among poisoned individuals, critical factors influencing mortality and survival include exposure to AlP (amount of aluminum consumed, tablet freshness) and the individual's personal characteristics; these should be considered essential parameters. The approach to managing AlPP should be based on a thorough understanding of its toxicological effects and associated clinical manifestations. Given that there is currently no specific antidote for AlP, treatment remains supportive and focuses on several key strategies. First, early gastrointestinal decontamination is crucial. Second, continuous hemodynamic monitoring is essential to manage cardiovascular instability, a common complication in severe cases. Ultimately, a holistic treatment approach that includes multidisciplinary collaboration among intensivists, cardiologists, and toxicologists can significantly improve patient outcomes and reduce mortality rates associated with AlPP. This review highlights various therapeutic strategies that have been explored to mitigate the effects of acute AlPP.

Gastric decontamination with vegetable oils, including coconut and olive oil, has shown promise in inhibiting the release of phosphine gas by forming a barrier that prevents further absorption of the toxin. Adjunctive therapies, such as magnesium sulfate and liothyronine, have been investigated for potential benefit. Magnesium sulfate may help correct hypomagnesemia and reduce arrhythmias, while liothyronine, a synthetic thyroid hormone, may protect mitochondrial function and enhance cellular metabolism. Advanced supportive treatments, including ECMO and CRRT, have emerged as critical interventions for patients with severe cardiopulmonary compromise. These modalities support organ function and provide a window for recovery by stabilizing hemodynamic parameters and correcting severe metabolic disturbances. High-dose insulin therapy has been associated with improved cardiac output and systemic perfusion, possibly due to its inotropic and vasodilatory effects. Red blood cell exchange has been considered in cases of significant hemolysis caused by oxidative stress, with the aim of removing damaged cells and replacing them with healthy cells.

Novel approaches, such as DHA administration, aim to bypass impaired cellular respiration by providing alternative substrates for ATP production, thereby sustaining vital cellular functions. Antioxidant therapies using agents such as NAC, coenzyme Q10, and vitamin E aim to mitigate phosphine-induced oxidative damage and potentially improve survival rates. Despite these varied interventions, AlPP management remains predominantly supportive. The effectiveness of many treatments is based on limited case reports and small studies, underscoring the need for more extensive, controlled clinical trials to establish standardized protocols. Early recognition and prompt initiation of combined therapeutic strategies are key to improving patient outcomes.

While significant progress has been made in exploring potential treatments for AlPP, there remains a crucial need for ongoing research to develop evidence-based guidelines. Multidisciplinary collaboration and increased awareness can enhance the management of this life-threatening condition and reduce its global impact.

Ethics

Acknowledgment

This study is the result of an in-house investigation and does not receive external financial support. The authors confirm that they have no conflicts of interest.

Footnotes

Authorship Contributions

Surgical and Medical Practices: S-A.B., A.B., H.H-A., M.B., S.H., A.J., O.M., M.A., Concept: A.B., O.M., M.A., Analysis or Interpretation: H.H-A., M.B., O.M., Writing: S-A.B., S.H., A.J.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Anger F, Paysant F, Brousse F, Le Normand I, Develay P, Gaillard Y, et al. Fatal aluminum phosphide poisoning. J Anal Toxicol. 2000;24:90-2.
2. Hekmatirad S, Abdollahi M. The possible link between aluminum phosphide poisoning and NLRP3 inflammasome. Medical Hypotheses. 2024;187:111352.
3. Aghebat-Bekheir S, Abdollahi M. Discovering the most impactful treatments for aluminum phosphide cardiotoxicity gleaned from systematic review of animal studies. Hum Exp Toxicol. 2024;43:9603271241290922.
4. Shumet A, Shiferaw N, Mekonnen D, Asemahagn MA. Trends and outcomes of acute poisoning in felege hiwot comprehensive specialized hospital medical intensive care units: retrospective study. Open Access Emerg Med. 2022;14:649-56.
5. Bumbrah GS, Krishan K, Kanchan T, Sharma M, Sodhi GS. Phosphide poisoning: a review of literature. Forensic Sci Int. 2012;214:1-6.
6. Gunnell D, Eddleston M, Phillips MR, Konradsen F. The global distribution of fatal pesticide self-poisoning: systematic review. BMC Public Health. 2007;7:357.
7. Soltaninejad K, Nelson LS, Bahreini SA, Shadnia S. Fatal aluminum phosphide poisoning in Tehran-Iran from 2007 to 2010. Indian J Med Sci. 2012;66:66-70.
8. Mehrpour O, Jafarzadeh M, Abdollahi M. A systematic review of aluminium phosphide poisoning. Arh Hig Rada Toksikol. 2012;63:61-73.

9. Gurjar M, Baronia AK, Azim A, Sharma K. Managing aluminum phosphide poisonings. J Emerg Trauma Shock. 2011;4:378-84.
10. Khan AM, Chaudhry H, Dagar M, Khatri CP, Amedu OS, Kumar U, et al. Aluminum phosphide poisoning: a cross-sectional study of clinical variables linked to inpatient mortality. SSRN Electron J. 2022.
11. Mehrpour O, Alfred S, Shadnia S, Keyler DE, Soltaninejad K, Chalaki N, et al. Hyperglycemia in acute aluminum phosphide poisoning as a potential prognostic factor. Hum Exp Toxicol. 2008;27:591-5.
12. Djurhuus R. Fumigation on bulk cargo ships: a chemical threat to seafarers. Int Marit Health. 2021;72:206-16.
13. Nasri Nasrabadi Z, Marashi SM. Comments on “a systematic review of aluminium phosphide poisoning”. Arh Hig Rada Toksikol. 2012;63:551.
14. Sanaei-Zadeh H. Response to "blood levels of methemoglobin in patients with aluminum phosphide poisoning and its correlation with patient's outcome". J Med Toxicol. 2012;8:86-7.
15. Sanaei-Zadeh H. Aluminum phosphide poisoning and development of hemolysis and methemoglobinemia. Indian J Crit Care Med. 2012;16:248-9.
16. Sanaei-Zadeh H, Marashi SM. Gastric decontamination in aluminium phosphide poisoning: a case against the use of water-based solutions. Arh Hig Rada Toksikol. 2016;67:364-65.
17. Goswami M, Bindal M, Sen P, Gupta SK, Avasthi R, Ram BK. Fat and oil inhibit phosphine release from aluminium phosphide--its clinical implication. Indian J Exp Biol. 1994;32:647-9.
18. Pajoumand A, Jalali N, Abdollah M, Shadnia S. Survival following severe aluminium phosphide poisoning. J Pharm Pract Res. 2002;32:297-9.
19. Agrawal VK, Bansal A, Singh RK, Kumawat BL, Mahajan P. Aluminum phosphide poisoning: possible role of supportive measures in the absence of specific antidote. Indian J Crit Care Med. 2015;19:109-12.
20. Altintop I, Tatli M. Successful management of zinc phosphide poisoning: possible benefit of virgin olive oil. Med Sci. 2017;6:537-40.
21. Shadnia S, Rahimi M, Pajoumand A, Rasouli MH, Abdollahi M. Successful treatment of acute aluminium phosphide poisoning: possible benefit of coconut oil. Hum Exp Toxicol. 2005;24:215-8.
22. Bajwa SJ, Bajwa SK, Kaur J, Singh K, Panda A. Management of celphos poisoning with a novel intervention: a ray of hope in the darkest of clouds. Anesth Essays Res. 2010;4:20-4.
23. Siwach SB, Singh P, Ahlawat S, Dua A, Sharma D. Serum & tissue magnesium content in patients of aluminium phosphide poisoning and critical evaluation of high dose magnesium sulphate therapy in reducing mortality. J Assoc Physicians India. 1994;42:107-10.
24. Siwach SB, Dua A, Sharma R, Sharma D, Mehla RK. Tissue magnesium content and histopathological changes in non-survivors of aluminium phosphide poisoning. J Assoc Physicians India. 1995;43:676-8.
25. Chugh SN, Kolley T, Kakkar R, Chugh K, Sharma A. A critical evaluation of anti-peroxidant effect of intravenous magnesium in acute aluminium phosphide poisoning. Magnes Res. 1997;10:225-30.
26. Abdolghaffari AH, Baghaei A, Solgi R, Gooshe M, Baeeri M, Navaei-Nigjeh M, et al. Molecular and biochemical evidences on the protective effects of triiodothyronine against phosphine-induced cardiac and mitochondrial toxicity. Life Sci. 2015;139:30-9.
27. Goharbari MH, Taghaddosinejad F, Arefi M, Sharifzadeh M, Mojtahedzadeh M, Nikfar S, et al. Therapeutic effects of oral liothyronine on aluminum phosphide poisoning as an adjuvant therapy: a clinical trial. Hum Exp Toxicol. 2018;37:107-17.
28. Mehrpour O, Farzaneh E, Abdollahi M. Successful treatment of aluminum phosphide poisoning with digoxin: a case report and review of literature. Int J Pharmacol. 2011;7:761-4.
29. Oghabian Z, Mehrpour O. Treatment of aluminium phosphide poisoning with a combination of intravenous glucagon, digoxin and antioxidant agents. Sultan Qaboos Univ Med J. 2016;16:e352-5.
30. Webb CA, Weyker PD, Flynn BC. Management of intra-aortic balloon pumps. Semin Cardiothorac Vasc Anesth. 2015;19:106-21.
31. Dukhin O, Bala D, Felker E, Golovina P, Tretyakova M, Haes B, et al. Case report: combination of veno-arterial extracorporeal membrane oxygenation and intra-aortic balloon pump in a young male patient with refractory cardiogenic shock due to aluminum phosphide poisoning. Front Cardiovasc Med. 2023;10:1226827.
32. Bagheri-Moghaddam A, Abbaspour H, Tajoddini S, Mohammadzadeh V, Moinipour A, Dadpour B. Using intra-aortic balloon pump for management of cardiogenic shock following aluminum phosphide poisoning; report of 3 cases. Emerg (Tehran). 2018;6:e3.
33. Mehrpour O, Asadi S, Yaghoubi MA, Azdaki N, Mahmoodabadi N, Javadmoosavi S. Cardiogenic shock due to aluminum phosphide poisoning treated with intra-aortic balloon pump: a report of two cases. Cardiovasc Toxicol. 2019;19:474-81.
34. Mehrpour O, Amouzeshi A, Dadpour B, Oghabian Z, Zamani N, Amini S, et al. Successful treatment of cardiogenic shock with an intraaortic balloon pump following aluminium phosphide poisoning. Arh Hig Rada Toksikol. 2014;65:121-6.
35. Siddaiah L, Adhyapak S, Jaydev S, Shetty G, Varghese K, Patil C, et al. Intra-aortic balloon pump in toxic myocarditis due to aluminum phosphide poisoning. J Med Toxicol. 2009;5:80-3.
36. Sauer CM, Yuh DD, Bonde P. Extracorporeal membrane oxygenation use has increased by 433% in adults in the United States from 2006 to 2011. ASAIO J. 2015;61:31-6.
37. Mohan B, Gupta V, Ralhan S, Gupta D, Puri S, Wander GS, et al. Role of extracorporeal membrane oxygenation in aluminum phosphide poisoning-induced reversible myocardial dysfunction: a novel therapeutic modality. J Emerg Med. 2015;49:651-6.
38. Sharma A, Sharma A, Acharya A, Aryal D, Rajbanshi BG, Bhattarai PR, et al. Extracorporeal membrane oxygenation in aluminum phosphide poisoning in Nepal: a case report. J Med Case Rep. 2018;12:1-5.
39. Hassanian-Moghaddam H, Zamani N, Rahimi M, Hajesmaeili M, Taherkhani M, Sadeghi R. Successful treatment of aluminium phosphide poisoning by extracorporeal membrane oxygenation. Basic Clin Pharmacol Toxicol. 2016;118:243-6.
40. Mohan B, Singh B, Gupta V, Ralhan S, Gupta D, Puri S, et al. Outcome of patients supported by extracorporeal membrane oxygenation for aluminum phosphide poisoning: an observational study. Indian Heart J. 2016;68:295-301.
41. Tandukar S, Palevsky PM. Continuous renal replacement therapy: who, when, why, and how. Chest. 2019;155:626-38.
42. Nasa P, Gupta A, Mangal K, Nagrani SK, Raina S, Yadav R. Use of continuous renal replacement therapy in acute aluminum phosphide poisoning: a novel therapy. Ren Fail. 2013;35:1170-2.
43. Prowle JR, Bellomo R. Continuous renal replacement therapy: recent advances and future research. Nat Rev Nephrol. 2010;6:521-9.
44. Bashardoust B, Farzaneh E, Habibzadeh A, Seyyed Sadeghi MS. Successful treatment of severe metabolic acidosis due to acute aluminum phosphide poisoning with peritoneal dialysis: a report of 2 cases. Iran J Kidney Dis. 2017;11:165-7.
45. Finkle SN. Should dialysis be offered in all cases of metformin-associated lactic acidosis? Crit Care. 2009;13:110.
46. Jadhav AP, Nusair MB, Ingole A, Alpert MA. Unresponsive ventricular tachycardia associated with aluminum phosphide poisoning. Am J Emerg Med. 2012;30:633.e3-5.

47. Açıkalın A, Dişel NR, Karakoç E, Matyar S, Sebe A. Successful treatment of aluminum phosphide poisoning with continuous veno-venous hemofiltration: a case report. J Emerg Med Case Rep. 2016;7:53-5.
48. Sedaghattalab M. Treatment of critical aluminum phosphide (rice tablet) poisoning with high-dose insulin: a case report. J Med Case Rep. 2022;16:192.
49. Hassanian-Moghaddam H, Zamani N. Therapeutic role of hyperinsulinemia/euglycemia in aluminum phosphide poisoning. Medicine (Baltimore). 2016;95:e4349.
50. Pannu AK, Bhalla A, Gantala J, Sharma N, Kumar S, Dhibar DP. Glucose-insulin-potassium infusion for the treatment of acute aluminum phosphide poisoning: an open-label pilot study. Clin Toxicol (Phila). 2020;58:1004-9.
51. Rahimi N, Abdolghaffari AH, Partoazar A, Javadian N, Dehpour T, Mani AR, et al. Fresh red blood cells transfusion protects against aluminum phosphide-induced metabolic acidosis and mortality in rats. PLoS One. 2018;13:e0193991.
52. Zamani N, Hassanian-Moghaddam H, Ebrahimi S. Whole blood exchange transfusion as a promising treatment of aluminium phosphide poisoning. Arh Hig Rada Toksikol. 2018;69:275-7.
53. Humayun M, Haider I, Badshah A, Subhan F. Protective role of G6PD deficiency in aluminium phosphide poisoning. J Coll Physicians Surg Pak. 2015;25 Suppl 1:S66-8.
54. Srinivas R, Agarwal R, Jairam A, Sakhuja V. Intravascular haemolysis due to glucose-6-phosphate dehydrogenase deficiency in a patient with aluminium phosphide poisoning. Emerg Med J. 2007;24:67-8.
55. Ortega MA, Fraile-Martinez O, García-Montero C, Callejón-Peláez E, Sáez MA, Álvarez-Mon MA, et al. A general overview on the hyperbaric oxygen therapy: applications, mechanisms and translational opportunities. Medicina (Kaunas). 2021;57:864.
56. Saidi H, Shokraneh F, Ghafouri HB, Shojaie S. Effects of hyperbaric oxygenation on survival time of aluminum phosphide intoxicated rats. J Res Med Sci. 2011;16:1306-12.
57. Hájek M, Chmelař D, Tlapák J, Rybárová V, Ondra P, Halouzka V. Accidental aluminum phosphide intoxication successfully treated with hyperbaric oxygen therapy: a case report. Toxics. 2024;12:272.
58. Niknahad H, Heidari R, Hashemi A, Jamshidzadeh A, Rashedinia M. Antidotal effect of dihydroxyacetone against phosphine poisoning in mice. J Biochem Mol Toxicol. 2021;35:e22897.
59. Oghabian Z, Ahmadi J, Pakravan S, Dabaghzadeh F, Heidari MR, Tajaddini S, et al. Successful treatment of aluminium phosphide poisoning by dihydroxyacetone: a two-case report study. J Clin Pharm Ther. 2020;45:1194-8.
60. Niknahad H, Heidari R, Jangjou A, Asghari V, Niknahad FM, Goudarzi F, et al. The therapeutic effect of a novel parenteral formulation of dihydroxyacetone in aluminum phosphide-intoxicated patients. Heliyon. 2023;9:e22165.
61. Bogle RG, Theron P, Brooks P, Dargan PI, Redhead J. Aluminium phosphide poisoning. Emerg Med J. 2006;23:e3.
62. Azad A, Lall SB, Mittra S. Effect of N-acetylcysteine and L-NAME on aluminium phosphide induced cardiovascular toxicity in rats. Acta Pharmacol Sin. 2001;22:298-304.
63. Marashi SM, Majidi M, Sadeghian M, Jafarzadeh M, Mohammadi S, Nasri-Nasrabadi Z. Protective role of coenzyme Q10 as a means of alleviating the toxicity of aluminum phosphide: an evidence-based review. Tzu Chi Medical Journal. 2015;27:7-9.
64. Keogh A, Fenton S, Leslie C, Aboyoun C, Macdonald P, Zhao YC, et al. Randomised double-blind, placebo-controlled trial of coenzyme Q, therapy in class II and III systolic heart failure. Heart Lung Circ. 2003;12:135-41.
65. Sander S, Coleman CI, Patel AA, Kluger J, White CM. The impact of coenzyme Q10 on systolic function in patients with chronic heart failure. J Card Fail. 2006;12:464-72.
66. Stafford RS, Radley DC. The underutilization of cardiac medications of proven benefit, 1990 to 2002. J Am Coll Cardiol. 2003;41:56-61.
67. Halvaei Z, Tehrani H, Soltaninejad K, Abdollahi M, Shadnia S. Vitamin E as a novel therapy in the treatment of acute aluminum phosphide poisoning. Turk J Med Sci. 2017;47:795-800.
68. Katwal S, Malbul K, Mandal SK, Kc S, Alam MZ, Karki P, et al. Successfully managed aluminum phosphide poisoning: a case report. Ann Med Surg (Lond). 2021;70:102868.
69. Shen Y, Jiang D, Wang T, Li M, Wang Y, Li J, et al. Blocking vicious cycle of cardiac dysfunction by external cardiopulmonary resuscitation is the key to rescue acute aluminium phosphide poisoning. Perfusion. 2023;38:214-9.
70. Kamaloddini MH, Zohoorian P, Foroughian M, Awli SH, Teimouri A. Successful treatment of acute aluminum phosphide poisoning: possible benefit of olive oil-a case report. Updates in Emergency Medicine. 2021;1:47-51.
71. Özkale M, Özkale Y, Kozanoglu İ. Role of automated red blood cell exchange in the treatment of aluminum phosphide poisoning: a case report and review of the literature. J Clin Apher. 2022;37:320-5.
72. Kamaloddini MH, Zohoorian P, Foroughian M, Hemmati Awli S, Teimouri A. Successful treatment of acute aluminum phosphide poisoning: possible benefit of olive oil-a case report. Updates Emerg Med. 2021;1:6-10.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.93585
Eurasian J Emerg Med. 2026;25: 342-50

Temporal and Meteorological Patterns of Assault-related Emergency Department Presentations

Abdullah Emin Gönülal[1], Secdegül Coşkun Yaş[2], Serdal Ateş[3], Cansu Uludağ Tunçel[4], Yahya Kemal Günaydın[3]

[1]Akyurt State Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[2]Gazi University Faculty of Medicine, Department of Emergency Medicine, Ankara, Türkiye
[3]University of Health Sciences Türkiye, Ankara Training and Research, Clinic of Emergency Medicine, Ankara, Türkiye
[4]Karabük Training and Research Hospital, Clinic of Emergency Medicine, Karabük, Türkiye

Abstract

Aim: Assault-related injuries constitute a significant burden on emergency departments and may be influenced by environmental factors. This study aimed to explore the association between meteorological parameters and assault-related emergency department presentations in an urban, upper-middle-income setting, and to describe temporal and clinical characteristics relevant to emergency service planning.

Materials and Methods: This retrospective observational study included adult patients presenting with assault-related injuries to a tertiary emergency department between January 1 and December 31, 2023. Demographic characteristics, injury mechanisms and types, clinical outcomes, and time of presentation were recorded. Meteorological data, including daily mean temperature, wind speed, and precipitation status, were obtained from official records. Associations between weather variables, temporal factors, and clinical outcomes were analyzed.

Results: A total of 4,042 patients were included: 2,606 (64.5%) were male, and the median age was 32 years. Nearly half of the presentations occurred between 16:00 and 23:59 (48.6%), with the highest frequency observed on Sundays (16.1%) and during the summer (31.4%). Physical assault was the predominant mechanism of injury (88.3%), and soft tissue injury was the most common injury type (78.9%). Assault-related presentations peaked when daily mean temperatures ranged between 20 °C and 30 °C, while wind speed and precipitation showed no significant pattern with presentation frequency. Stab/cut-related and firearm-related injuries were significantly more frequent among patients who required hospitalization, who had life-threatening conditions, or who required surgery ($p<0.001$).

Conclusion: Assault-related emergency department presentations demonstrate distinct temporal and environmental patterns in an upper-middle-income urban setting. An increased frequency of presentations within specific temperature ranges and during predictable time periods highlights the importance of incorporating environmental and temporal factors into emergency department resource planning and violence prevention strategies.

Keywords: Assault, emergency service, violence, weather, temperature

Introduction

Injuries are among the leading causes of death and disability worldwide and impose a substantial burden on emergency care (1,2). The World Health Organization classifies injuries broadly as unintentional and intentional (violence-related) injuries (3). The high volume of injury-related emergency department visits underscores the need to address not only clinical management but also resource planning and safety considerations.

Violence-related injuries are predominantly caused by blunt or penetrating trauma mechanisms. Blunt trauma typically results from physical assault, whereas penetrating trauma is caused by sharp objects or firearms (4). The clinical management and

Corresponding Author: Abdullah Emin Gönülal MD, Akyurt State Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
E-mail: abdullahemingonulal@gmail.com **ORCID ID:** orcid.org/0009-0002-5840-836X

Cite this article as: Gönülal AE, Yaş SC, Ateş S, Uludağ Tunçel C, Günaydın YK. Temporal and meteorological patterns of assault-related emergency department presentations. Eurasian J Emerg Med. 2026;25: 342-50.

Received: 12.02.2026
Accepted: 02.06.2026
Published: 26.06.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

outcomes associated with these mechanisms differ, necessitating distinct diagnostic and therapeutic approaches (5). Beyond clinical management, however, understanding the factors that precipitate human aggression is essential for effective prevention and emergency preparedness.

Weather represents a key environmental factor that may influence daily human behavior and modify the incidence of aggression and trauma. Previous studies have demonstrated a positive association between increasing ambient temperatures and the frequency of aggression- and trauma-related cases (6,7). In addition to temperature, the potential effects of other meteorological variables, including wind speed, precipitation, and lunar phases, on trauma incidence have been explored (7). The observed increase in emergency department visits during hot summer months and on weekends further suggests that environmental influences on human behavior should be taken into account in emergency healthcare planning (8).

The distribution of injury cases varies across global, national, and local contexts, underscoring the influence of sociocultural, environmental, and climatic factors (6,9). This study aimed to examine associations between weather parameters and the frequency of assault-related injury presentations to an emergency department in an upper-middle-income urban setting and to highlight the implications of environmental conditions for forecasting emergency department demand and service planning.

Materials and Methods

This retrospective observational study was conducted among patients presenting to the emergency department of a tertiary care hospital. The study included patients aged 18 years and older who presented to the emergency department due to assault-related injuries between January 1, 2023, and December 31, 2023. This study was conducted in accordance with the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee of University of Health Sciences Türkiye, Ankara Training and Research (approval no: E-24-109, date: 10. 05.2024). Due to the retrospective design, informed consent was waived. Patients with inaccessible medical records, pregnant patients, and those injured due to other forensic causes (suicide attempts, drug intoxication, occupational accidents, or traffic-related accidents) were excluded.

Patient data were extracted from the hospital's electronic health record system. Cases were identified using a combination of ICD-10 codes (including W50, W51, W26, Z04.5, and Y28, where applicable), physician documentation in emergency department records, and forensic case classification within the electronic health record system. Recorded variables included demographic characteristics (age, gender), time of presentation (hour, day, and month), mechanism of injury, type of injury, consultation requirement, need for surgical intervention, hospital admission, and 24-hour mortality. A life-threatening condition was defined as any clinical situation requiring immediate intervention to prevent death, including severe hemorrhage, intracranial injury, or hemodynamic instability. Emergency surgical intervention was defined as any operative procedure performed within the first 24 hours following admission for the injury. Patients were categorized into three groups according to the time of presentation: 00:00-07:59, 08:00-15:59, and 16:00-23:59. Assault mechanisms were classified as physical assault (assault by bodily force), stab/cut injuries, firearm injuries, sexual assault, chemical irritant exposure (pepper spray), and burn injury. Injury types were categorized as soft tissue trauma, maxillofacial trauma, extremity fractures, complex wounds, thoracic trauma, abdominal trauma, intracranial trauma, and vertebral trauma. Meteorological data for the study location, including daily mean temperature, wind speed, and precipitation status, were obtained from publicly available aggregated datasets (WeatherSpark), which compile data from official meteorological stations, and entered into the study database. Precipitation was categorized as present or absent based on daily records.

Statistical Analysis

Statistical analyses were performed using Jamovi software (version 2.5.7). Categorical variables were summarized as frequencies and percentages, while continuous variables were presented as medians and interquartile ranges (25^{th}-75^{th} percentiles). Data distribution was assessed using the Kolmogorov-Smirnov test. The Mann-Whitney U test was used to compare continuous variables between groups. For categorical variables, the chi-square test was employed. Bonferroni correction was applied for multiple comparisons, and a p-value<0.05 was considered statistically significant. Monthly case counts were normalized by calculating the mean daily case count to account for differences in the number of days per month.

Results

A total of 4,042 patients with assault-related injuries who presented to the emergency department over a one-year period were included in the study. Of these, 64.5% were male (n=2.606), and the median age was 32 years (interquartile range 24-42). Regarding the time of presentation, the largest proportion of patients presented between 16:00 and 23:59 (n=1.964; 48.6%). Presentations were most frequent on Sundays (n=652; 16.1%) and during the summer season (n=1.271; 31.4%). Monthly data

were presented as mean daily case counts (cases/day) to avoid bias due to differences in the number of days across months. As illustrated in Figure 1, the highest presentation rates were observed during the summer months, peaking in June and remaining elevated through July and August. During the study period, the median daily temperature was 15 °C (range: -6 to 35 °C). The relationship between temperature variation and visit frequency is illustrated in Figure 2, while Figure 3 demonstrates the association between wind speed and the number of visits. Precipitation did not demonstrate a statistically significant association with presentation frequency. Baseline characteristics of the study population are summarized in Table 1.

Among patients presenting after assault-related incidents, 102 patients (2.5%) were identified as having a life-threatening condition, and 126 patients (3.1%) needed emergency surgical intervention. Evaluation of emergency department outcomes showed that 3,817 patients (94.2%) were discharged; 134 (3.3%) were admitted to hospital wards; 60 (1.5%) were admitted to the intensive care unit; and 31 (0.8%) were referred to another medical center. Assessment of 24-hour mortality revealed that 9 patients (0.2%) died.

When assault types were analyzed, physical assault was the most common mechanism, identified in 3,568 patients (88.3%). The most frequently observed injury type was soft tissue injury (78.9%). A total of 1,031 patients (25.5%) required a consultation. Injury types and consultation data are presented in Table 2.

Pairwise comparisons were performed with respect to hospital admission, presence of life-threatening conditions, and need for surgical intervention (Table 3). All clinical outcomes were significantly more frequent among male patients than among female patients ($p<0.001$). In terms of injury mechanisms, stab/cut injuries and firearm injuries were significantly more frequent among patients who required hospitalization, had life-threatening conditions, or needed surgical intervention compared with those who did not ($p<0.001$). Post-hoc analyses demonstrated that, for hospitalization and emergency surgery, all pairwise comparisons between trauma mechanisms were statistically significant after Bonferroni correction (all adjusted p-values <0.001). For life-threatening injuries, significant differences were observed between physical assault and both stab-cut and firearm injuries, whereas no significant difference was detected between stab-cut and firearm injuries.

Regarding the time of presentation, patients presenting during nighttime hours (00:00-07:59) exhibited higher rates of all three clinical outcomes, whereas patients presenting during daytime hours (08:00-15:59) exhibited lower rates; this difference was statistically significant for hospital admission ($p=0.011$) and the presence of life-threatening conditions ($p=0.026$). Post-hoc analyses revealed that statistically significant differences were observed only between the 00:00-07:59 and 08:00-15:59 intervals for both hospital admission and presence of life-threatening conditions (adjusted $p=0.018$ and $p=0.024$, respectively). No significant pairwise differences were detected between the other time intervals for either outcome.

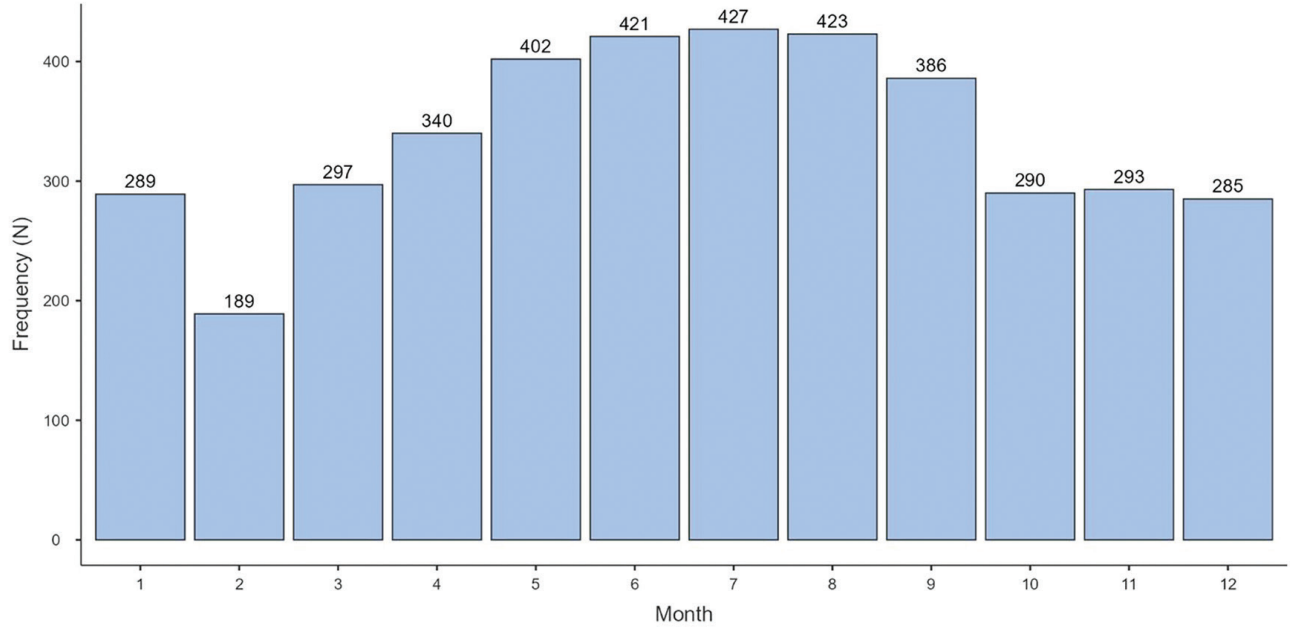


Figure 1. Monthly distribution of assault-related emergency department presentations expressed as mean daily case counts (cases/day)

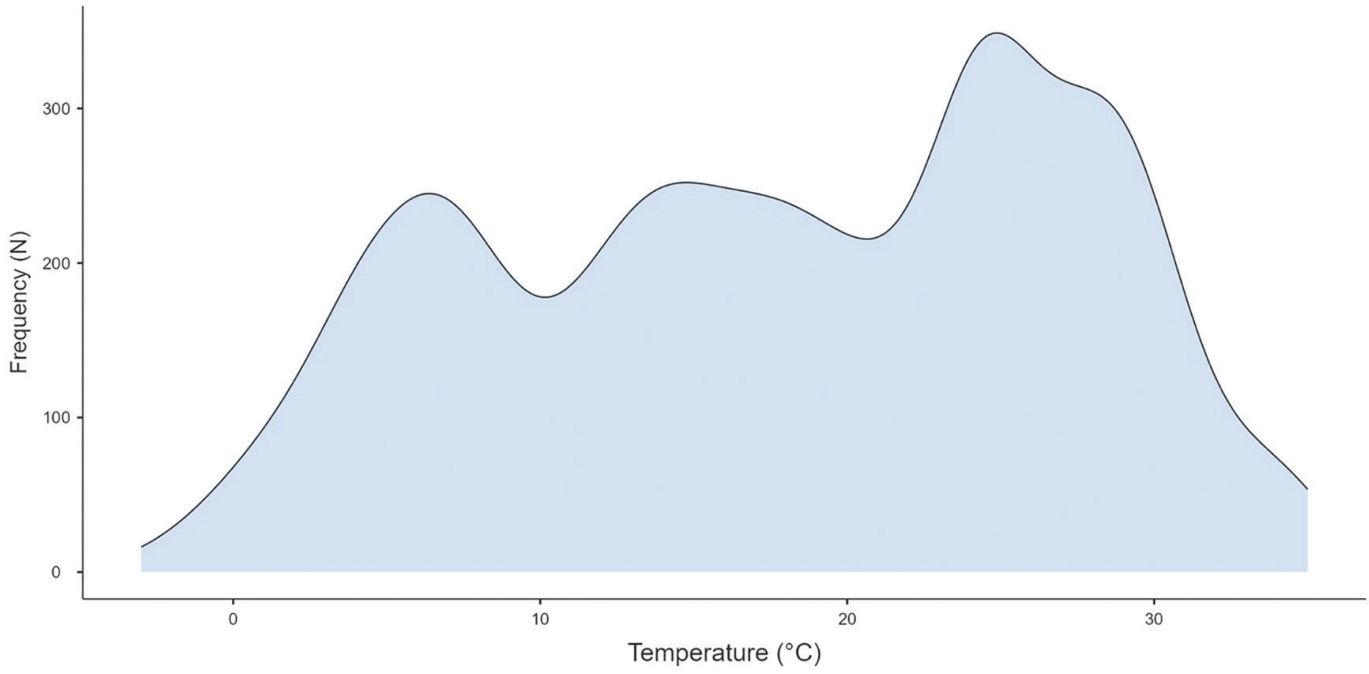


Figure 2. Frequency of assault-related injury presentations by the mean daily temperature on the day of presentation (°C). The plot is a histogram or area-style display; no smoothing was applied (raw counts per temperature bin)

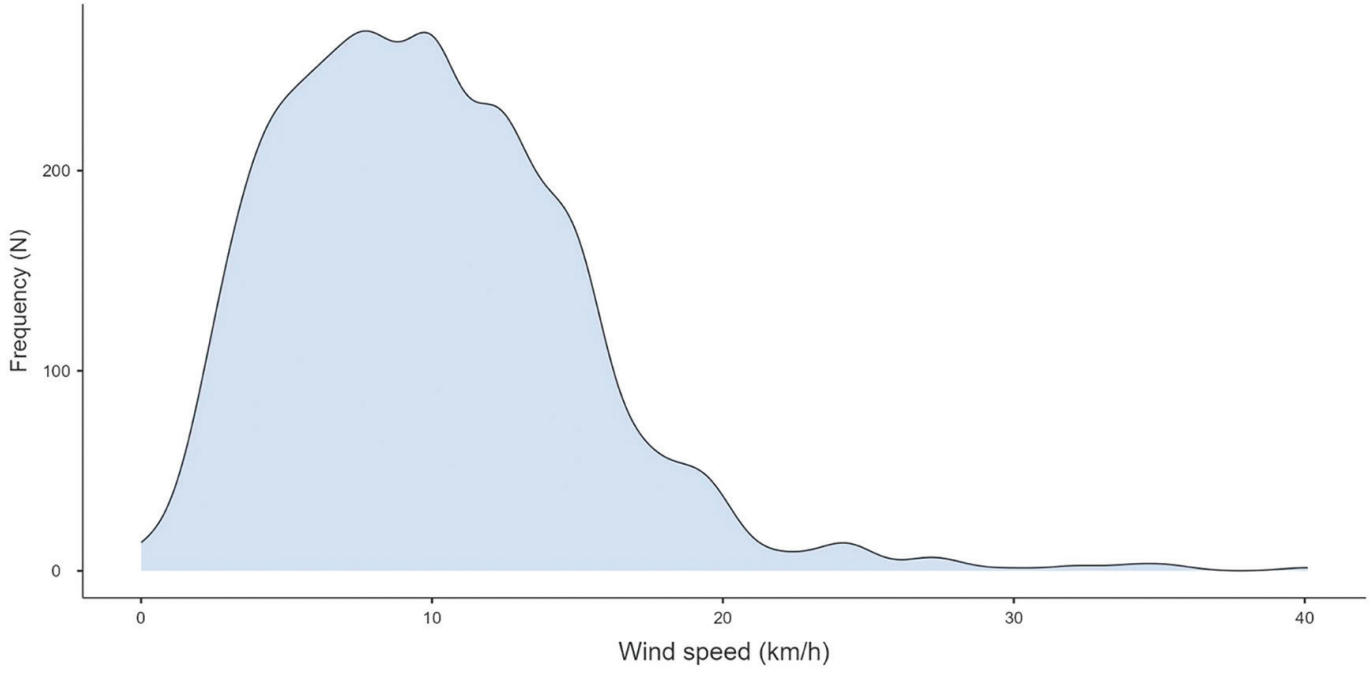


Figure 3. Frequency of assault-related injury presentations by the mean daily wind speed on the day of presentation (km/h). The plot is a histogram/area display; no smoothing was applied (raw counts by wind speed bins)

Table 1. Demographic characteristics, clinical outcomes, and temporal distribution of patients presenting with assault-related injuries	
	Total n= 4042
Male, gender, n (%)	2606 (64.5)
Age, median (IQR 25-75)	32 (24-42)
Temperature at presentation (°C), median (IQR 25-75)	15 (7-22)
Wind speed (km/h), median (IQR 25-75)	9.5 (6.0-13.0)
Patients with life-threatening injury, n (%)	102 (2.5)
Patients requiring surgery, n (%)	126 (3.1)
Emergency department outcome, n (%)	
Discharged	3817 (94.2)
Admitted to ward	134 (3.3)
Admitted to intensive care unit	60 (1.5)
Transferred to another center	31 (0.8)
24-hour mortality, n (%)	9 (0.2)
Time of presentation, n (%)	
00:00-08:00	1018 (25.2)
08:00-16:00	1060 (26.2)
16:00-00:00	1964 (48.6)
Day of presentation, n (%)	
Monday	635 (15.7)
Tuesday	537 (13.3)
Wednesday	626 (15.5)
Thursday	546 (13.5)
Friday	500 (12.4)
Saturday	546 (13.5)
Sunday	652 (16.1)
Season of presentation, n (%)	
Winter	763 (18.9)
Spring	1039 (25.7)
Summer	1271 (31.4)
Autumn	969 (24.0)
Precipitation status, n (%)	
Snow	123 (3.0)
Rain	1197 (29.6)
No precipitation	2722 (67.3)
IQR: Interquartile range	

Table 2. Distribution of injury mechanisms, injury types, and consultation requirements	
	Total n (%) n= 4042
Mechanism of injury	
Physical assault	3568 (88.3)
Stab/cut injury	331 (8.2)
Firearm injury	121 (3)
Pepper spray assault	10 (0.2)
Burn injury	8 (0.2)
Sexual assault	4 (0.1)
Type of injury	
Soft tissue injury	3190 (78.9)
Maxillofacial trauma	416 (10.3)
Extremity fracture	177 (4.4)
Complicated wound	137 (3.4)
Thoracic trauma	68 (1.7)
Abdominal trauma	49 (1.2)
Intracranial injury	34 (0.8)
Vertebral trauma	20 (0.5)
Consultation status	
Patients requiring consultation	1031 (25.5)
Otorhinolaryngology	413 (10.2)
Orthopedics and traumatology	278 (6.9)
Plastic and reconstructive surgery	228 (5.6)
Ophthalmology	227 (5.6)
Neurosurgery	75 (1.9)
General surgery	63 (1.6)
Cardiovascular surgery	57 (1.4)
Obstetrics and gynecology	15 (0.4)
Urology	15 (0.4)
Thoracic surgery	13 (0.3)
Other	8 (0.2)

Table 3. Comparison of demographic characteristics, injury mechanisms, and environmental factors according to clinical outcomes

Variables	Hospital admission		p	Life-threatening condition		p	Need for surgery		p
	Present (n=225)	Absent (n=3817)		Present (n=102)	Absent (n=3940)		Present (n=126)	Absent (n=3916)	
Age, years, median (IQR)	29 (23-41)	32 (24-42)	0.097[1]	28 (23-44)	32 (24-42)	0.198[1]	29 (23-41)	32 (24-42)	0.128[1]
Male, n (%)	200 (88.9)	2406 (63)	**<0.001[2]**	93 (91,2)	2513 (63.8)	**<0.001[2]**	113 (89.7)	2493 (63.7)	**<0.001[2]**
Mechanism of injury, n (%)									
Physical assault	81 (2.3)	3487 (97.7)	**<0.001[2]**	32 (0.9)	3536 (99.1)	**<0.001[2]**	41 (1.1)	3527 (98.9)	**<0.001[2]**
Stab/cut injury	77 (23.3)	254 (76.7)		45 (13.6)	286 (86.4)		46 (13.9)	285 (86.1)	
Firearm injury	66 (54.5)	55 (45.5)		24 (19.8)	97 (80.2)		37 (30.6)	84 (69.4)	

Table 3. Continued

Variables	Hospital admission		P	Life-threatening condition		P	Need for surgery		P
	Present (n=225)	Absent (n=3817)		Present (n=102)	Absent (n=3940)		Present (n=126)	Absent (n=3916)	
Time-of-day interval, n (%)									
00:00–07:59	75 (33.3)	943 (24.7)		36 (35.3)	982 (24.9)		43 (34.1)	975 (24.9)	
08:00–15:59	48 (21.3)	1012 (26.5)	**0.011**[2]	18 (17.6)	1042 (26.4)	**0.026**[2]	29 (23)	1031 (26.3)	0.063[2]
16:00–23:59	102 (45.4)	1862 (48.8)		48 (47.1)	1916 (48.7)		54 (42.9)	1910 (48.8)	
Precipitation, present, n (%)	72 (32.0)	1248 (32.7)	0.829[2]	34 (33.3)	1286 (32.6)	0.883[2]	39 (31.0)	1281 (32.7)	0.678[2]
Wind speed (km/h), median (IQR)	9.3 (6.0-12.5)	9.5 (6.0-13.0)	0.499[1]	10.0 (6.5-13.3)	9.5 (6.0-13.0)	0.385[1]	9.0 (5.9-12.0)	9.5 (6.0-13.0)	0.171[1]
Temperature (°C), median (IQR)	15 (7-21)	15 (7-22)	0.993[1]	17 (10-22)	15 (7-22)	0.098[1]	13 (5-21)	15 (7-22)	0.202[1]

IQR: Interquartile range. [1]Mann Whitney U test, [2]Chi-square test

Discussion

This study describes the profile of 4,042 patients who presented to the emergency department with assault-related injuries over a one-year period. Nearly half of all presentations occurred during the evening hours (16:00-23:59), with a noticeable increase on Sundays and during the summer season. Regarding injury mechanism, physical assault was by far the most common cause (88.3%), and soft tissue injuries predominated among injury types (78.9%). Taken together, our findings suggest a temporal clustering of assault-related presentations and a temperature-related pattern. However, these results do not establish a clear independent association. Violence remains a growing public health concern worldwide, including in our country. In particular, the increasing frequency of violence in developing and upper-middle-income countries contributes substantially to emergency department crowding and places emergency physicians in a critical position for the initial management of these patients (10). Emergency departments serve as the primary points of contact for victims of assault, where rapid assessment, acute management, and decisions regarding further care are initiated. Therefore, examining the clinical characteristics of assault-related injuries, injury mechanisms, and their relationship with environmental factors may provide valuable insights for healthcare planning and resource allocation.

The median age of patients presenting with assault-related injuries was 32 years; 64.5% were male. These findings are consistent with previous studies evaluating forensic and assault-related cases, which have similarly reported a predominance of young male victims (11,12). The overrepresentation of males and younger age groups among victims of violence is a well-established finding in the literature.

Temporal analyses in our cohort demonstrated that assault-related presentations were most frequent during the summer months (31.4%) and on Sundays (16.1%). Atherton et al. (13) reported a higher volume of trauma-related emergency department visits during summer periods. Similarly, Lemon et al. (7) observed that assault cases were more likely to present during weekends, and Khurana et al. (8) reported a significant increase in violent incidents during holiday periods. This temporal increase may be explained by increased social interaction, outdoor activities, and population mobility on weekends and during warmer seasons, which may collectively increase the risk of interpersonal conflict.

In our study, the most common mechanisms of assault-related injuries were physical assaults (88.3%), followed by stab or cut injuries (8.2%) and firearm injuries (3.0%). Similarly, in the study by Demircan et al. (12), which evaluated all forensic cases, physical assault was reported as the most frequent mechanism among violence-related injuries, followed by stab/cut injuries and gunshot injuries. Regarding injury type, soft tissue injury was the most common finding in our cohort (78.9%). Korkmaz et al. (14) also reported soft tissue injuries as the most frequent injury type among forensic cases in their study.

Several studies in the literature have examined the relationship between trauma and assault-related events and weather conditions (15,16). Ho et al. (11) reported a positive association between increasing temperature and trauma-related emergency department presentations. Michel et al. (17) similarly demonstrated that increases in daily maximum temperature were

associated with higher rates of violent crime. Livingston et al. (18) showed that pediatric trauma-related orthopedic consultations and emergency department visits increased by 1% for each 1 °C rise in temperature, while increasing precipitation was associated with a decrease in presentations. The capital city where this study was conducted has a climate characterized by four distinct seasons. In our cohort, the highest number of assault-related presentations was observed when daily mean temperatures ranged between 20 °C and 30 °C. It has been suggested that higher temperatures may increase outdoor activities and social interactions, thereby expanding opportunities for violent encounters; our findings are consistent with this interpretation. However, this observation should be interpreted as a descriptive pattern rather than a causal relationship. In contrast, wind speed did not demonstrate a linear relationship with the number of assault-related presentations in our study. Similarly, previous studies have not identified a significant association between wind speed and violence- or trauma-related emergency department visits (17,19,20). Overall, meteorological variables may be associated with variations in presentation frequency; however, their independent effect could not be established in this study.

When injury mechanisms were evaluated in relation to clinical severity, the proportion of life-threatening conditions was low among physical assault cases but was substantially higher among firearm and stab/cut injuries. In our study, life-threatening conditions were identified in 0.9% of physical assault cases, 13.6% of stab/cut injuries, and 19.8% of firearm injuries. Şenavcı and Erel (21) reported mortality rates of 0.04% for physical assault, 33.1% for sharp-object injuries, and 29.1% for firearm-related injuries, while Karasu et al. (22) reported rates of 5.5%, 48.4%, and 33.7%, respectively. These findings indicate that morbidity and mortality vary markedly according to the mechanism of violence, with penetrative injuries carrying substantially higher risk.

In our cohort, 3.1% of patients presenting with assault-related injuries required emergency surgical intervention. When cases were stratified by injury mechanism, emergency surgery was performed in 1.1% of physical assault cases, 13.9% of stab/cut injuries, and 30.6% of firearm injuries. Hazra et al. (23), in their study of assault patients in India, reported an emergency surgical intervention rate of 15.7%. Although the overall number of assault-related presentations in our study was high, the relatively low rates of hospital admission and surgical intervention suggest that a substantial proportion of patients presented with less severe injuries. This interpretation is further supported by the predominance of isolated soft tissue injuries in our cohort.

The findings of this study support strengthening trauma care capacity during evening and night shifts, including availability of experienced staff, timely access to surgical and imaging services, and adequate security measures. Planned reinforcement of staffing and resource allocation may be particularly beneficial on Sundays and during the summer. From a public health and municipal perspective, preventive field interventions during weekends, especially on warmer days, may be considered. Overall, our results indicate that timing, injury mechanisms, and seasonal factors should be taken into account in emergency department service planning.

Study Limitations

This study has several limitations. First, due to its single-center, retrospective design, the generalizability of the findings is limited, and causal inferences cannot be made. Second, the primary outcome measure was the number of emergency department presentations, which does not reflect population-based incidence and does not capture violent events that did not result in healthcare utilization. In addition, the analysis was primarily descriptive, and no multivariable adjustment was performed for temporal variables such as time of day, day of the week, and season. Third, meteorological data were derived from city-level averages and may not accurately reflect individual exposure, microclimatic variations, or potential discrepancies between the time and location of injury and the time of hospital presentation. Fourth, the study period was limited to one year; therefore, interannual climate variability and long-term trends could not be assessed. Furthermore, the observed association with temperature may reflect increased social interaction rather than a direct environmental effect. For these reasons, the findings should be interpreted as having value for planning and forecasting, and confirmation through multicenter studies with longer follow-up and enriched variable sets is warranted.

Conclusion

In conclusion, this study characterizes the clinical features, injury mechanisms, and temporal distribution of assault-related emergency department presentations in an upper-middle-income urban setting. An increase in presentations was observed within specific temperature ranges; however, this finding should be interpreted as a descriptive pattern rather than a confirmed independent association.

These findings may contribute to emergency department resource planning and to the development of preventive strategies. Further studies using multivariable analytical approaches are needed to clarify the relationship between environmental factors and assault-related injuries.

Ethics

Ethics Committee Approval: This study was conducted in accordance with the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee of University of Health Sciences Türkiye, Ankara Training and Research (approval no: E-24-109, date: 10. 05.2024).

Informed Consent: Due to the retrospective design, informed consent was waived.

Data Availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Footnotes

Authorship Contributions

Concept: A.E.G., S.C.Y., S.A., C.U.T., Design: A.E.G., S.C.Y., C.U.T., Y.K.G., Data Collection or Processing: A.E.G., S.A., C.U.T., Analysis or Interpretation: S.C.Y., S.A., Literature Search: A.E.G., Writing: A.E.G., S.C.Y., Y.K.G.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Peterson C, Miller GF, Barnett SBL, Florence C. Economic Cost of Injury - United States, 2019. MMWR Morb Mortal Wkly Rep. 2021;70:1655-9.
2. Haagsma JA, Graetz N, Bolliger I, Naghavi M, Higashi H, Mullany EC, et al. The global burden of injury: incidence, mortality, disability-adjusted life years and time trends from the Global Burden of Disease study 2013. Inj Prev. 2016;22:3-18.
3. Organization WH. Preventing injuries and violence: an overview. In: Organization WH, editor. 2022. Avaible from: https://iris.who.int/server/api/core/bitstreams/61d409eb-68c6-4e20-8d5b-a1ce82819b83/content
4. Varma D, Brown P, Clements W. Importance of the mechanism of injury in trauma radiology decision-making. Korean J Radiol. 2023;24:522-8.
5. Greaves I, Porter K, Garner J. Trauma care manual. CRC Press. 2021.
6. Bushman BJ, Wang MC, Anderson CA. Is the curve relating temperature to aggression linear or curvilinear? Assaults and temperature in minneapolis reexamined. J Pers Soc Psychol. 2005;89:62-6.
7. Lemon DJ, Partridge R; Pan-Dorset Cardiff Model team. Is weather related to the number of assaults seen at emergency departments? Injury. 2017;48:2438-42.
8. Khurana B, Prakash J, Loder RT. Assault related injury visits in US emergency departments: an analysis by weekday, month and weekday-by-month. Chronobiol Int. 2022;39:1068-77.
9. Sakran JV, Greer SE, Werlin E, McCunn M. Care of the injured worldwide: trauma still the neglected disease of modern society. Scand J Trauma Resusc Emerg Med. 2012;20:64.
10. Mercy JA, Krug EG, Dahlberg LL, Zwi AB. Violence and health: the United States in a global perspective. Am J Public Health. 2003;93:256-61.
11. Ho VP, Towe CW, Chan J, Barie PS. How's the weather? Relationship between weather and trauma admissions at a Level I Trauma Center. World J Surg. 2015;39:934-9.
12. Demircan A, Keleş A, Guerbuez N, Bildik F, Aygencel ŞG, Doğan NÖ, et al. Forensic emergency medicine-six-year experience of 13823 cases in a university emergency department. Turk J Med Sci. 2008;38:567-75.
13. Atherton WG, Harper WM, Abrams KR. A year's trauma admissions and the effect of the weather. Injury. 2005;36:40-6.
14. Korkmaz T, Kahramansoy N, Erkol Z, Sarıçil F, Kılıç A. Acil servise başvuran adli olguların ve düzenlenen adli raporların değerlendirilmesi [Evaluation of the forensic patients presenting to the emergency department and legal reports]. Med Bull Haseki. 2012;50:14-20. Turkish.
15. Lin LW, Lin HY, Hsu CY, Rau HH, Chen PL. Effect of weather and time on trauma events determined using emergency medical service registry data. Injury. 2015;46:1814-20.
16. Kim SH, Sul YH, Kim JS. Effect of weather variables on the incidence of trauma: a retrospective study at a single tertiary hospital center for 4 years. Medicine (Baltimore). 2021;100:e27623.
17. Michel SJ, Wang H, Selvarajah S, Canner JK, Murrill M, Chi A, et al. Investigating the relationship between weather and violence in Baltimore, Maryland, USA. Injury. 2016;47:272-6.
18. Livingston KS, Miller PE, Lierhaus A, Matheney TH, Mahan ST. Does weather matter? The effect of weather patterns and temporal factors on pediatric orthopedic trauma volume. Open Orthop J. 2016;10:550-8.
19. O'Connor ZJ, Helmer SD, Yates CL, Ward JG, Khandelwal A, Haan JM. An evaluation of the effect of wind speed on trauma admissions and injury severity at a midwest level I trauma center. American Surgeon. 2012;78:614-6.
20. Rising WR, O'Daniel JA, Roberts CS. Correlating weather and trauma admissions at a level I trauma center. J Trauma. 2006;60:1096-100.
21. Şenavcı M, Erel Ö. Travma skorlama sistemlerinin adli raporlarda yaşami tehlikeye sokan yaralanmanın şiddetini ayırt etmede kullanılabilirliği. Bull Leg Med. 2022;27:17-23.
22. Karasu M, Isır AB, Aydın N, Dülger HE. Gaziantep Üniversitesi Tıp Fakültesi Adli Tıp Anabilim Dalınca 1998-2005 yılları arasında düzenlenen adli raporların değerlendirilmesi. Gaziantep Medical Journal. 2009;15:10-5.
23. Hazra D, Nekkanti AC, Abhilash KPP. Injury pattern and outcome of assault victims: an emergency department perspective. Arch Trauma Res. 2020;9:154-9.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.57804
Eurasian J Emerg Med. 2026;25: 351-9

The Effect of Ramadan Fasting on Biliary System Diseases: A Prospective Analysis of Emergency Department Admissions

Levent Şahin[1], Merve Ayyıldız Akın[2], Taner Güven[3]

[1]Kafkas University Faculty of Medicine, Department of Emergency Medicine, Kars, Türkiye
[2]Kafkas University Faculty of Veterinary Medicine, Department of Veterinary Biostatistics, Kars, Türkiye
[3]Malatya Training and Research Hospital, Clinic of Emergency Medicine, Malatya, Türkiye

Abstract

Aim: Prolonged fasting impacts chronic diseases. This study aimed to describe the clinical, laboratory, radiological, and disposition characteristics of fasting patients with known biliary system disease who presented to the emergency department during Ramadan.

Materials and Methods: In this prospective descriptive study, we included fasting patients with known biliary diseases who presented to the emergency department during Ramadan with acute right upper quadrant pain, nausea, and/or vomiting. We recorded demographics (age, gender), laboratory parameters, and hospitalization/discharge status. Patients were categorized into four groups based on radiological imaging: normal findings, common bile duct (CBD) dilatation, cholelithiasis, and biliary sludge.

Results: The study included 37 patients. The most frequent pathological finding on radiological imaging was CBD dilatation (27%). Aspartate transaminase levels were significantly higher in females than in males ($p<0.05$). Furthermore, total and direct bilirubin levels were significantly elevated in patients with abnormal imaging findings and in those requiring hospitalization or referral ($p<0.001$). The hospitalization rate was also significantly higher for female patients ($p=0.007$).

Conclusion: Consuming high-calorie meals after prolonged Ramadan fasting may trigger symptomatic exacerbations in patients with biliary tract pathology. Elevated C-reactive protein and bilirubin levels were correlated with disease severity and the need for hospitalization. Patients with a history of biliary disease should consult their physician before fasting. Because the study was descriptive and lacked a control group, no definitive causal relationship can be established.

Keywords: Fasting, biliary system, starvation, Ramadan

Introduction

Intermittent fasting is a dietary approach involving cycles of fasting and non-fasting periods (1). The most common and popular form is the regimen that involves no calorie intake for two-thirds of the day (2). In our country, healthy Muslims practice this dietary pattern as a form of worship during the month of Ramadan. In the Islamic religion, fasting involves abstaining from eating and drinking from dawn until sunset (3). Therefore, fasting for 30 consecutive days during Ramadan can be considered a form of intermittent fasting. Numerous studies have been conducted to examine the course of existing diseases and investigate the changes that occur in people fasting during this month (4-6). Studies investigating the effects of prolonged hunger on human metabolism have focused on the positive effects of fasting on conditions such as obesity, hypertension, cardiovascular diseases, and diabetes (7,8). It has been observed that studies have prioritized issues such as dehydration caused by hunger and thirst, and the consequent renal colic and renal failure (9,10). Upon conducting a detailed literature review, we found a limited number of studies examining the effect of fasting on biliary system diseases (4). It has been reported that prolonged hunger and thirst tend to cause a decreasing trend in liver function enzymes and may trigger cholecystitis (11,12). Since less food is consumed during Ramadan, bile flow slows. This deceleration can cause concentrated bile to crystallize and

Corresponding Author: Levent Şahin MD, Kafkas University Faculty of Medicine, Department of Emergency Medicine, Kars, Türkiye
E-mail: levsahin44@gmail.com **ORCID ID:** orcid.org/0000-0003-0193-4393
Cite this article as: Şahin L, Ayyıldız Akın M, Güven T. The effect of ramadan fasting on biliary system diseases: a prospective analysis of emergency department admissions. Eurasian J Emerg Med. 2026;25: 351-9.

Received: 16.02.2026
Accepted: 08.06.2026
Published: 26.06.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

form stones. Prolonged fasting may produce negative outcomes in patients with pre-existing biliary problems (13). This study aimed to describe the clinical, laboratory, radiological, and disposition characteristics of fasting patients with known biliary system disease presenting to the emergency department during Ramadan.

Materials and Methods

This study was conducted prospectively in the emergency department of a tertiary hospital Kafkas University Health Practice and Research Hospital between March 11, 2024, and April 9, 2024. The study was initiated following the approval of the Kafkas Medical Faculty Ethics Committee (decision number: 02, date: 27.02.2024).

Study Design

Patients who were fasting and presented to the emergency department during Ramadan with gastrointestinal complaints, such as acute-onset right upper-quadrant pain were recorded. Among these patients, those who underwent blood parameter analysis and abdominal computed tomography (CT), ultrasonography (USG), or magnetic resonance cholangiopancreatography (MRCP) imaging were selected. Imaging was performed at the discretion of the treating emergency physician based on clinical presentation, symptom severity, and standard institutional protocols for the evaluation of acute right upper quadrant pain. No standardized imaging protocol specific to this study was applied beyond routine emergency department practice. Patients with no pathology in at least one of the imaging methods were classified as "normal"; those with "stones in the gallbladder" were classified as "cholelithiasis"; those with ductal dilatation due to stones were classified as "dilated common bile duct (CBD)"; and those with "biliary sludge" were classified as "biliary sludge." After obtaining informed consent from the patients to participate in the study, the investigators also recorded their age and gender. The results of "1 tube of hemogram and 1 tube of EDTA biochemistry" tests [white blood cell (WBC); liver function tests: alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transferase (GGT), lactate dehydrogenase (LDH); and C-reactive protein (CRP)], which are routinely performed in the emergency department, were added to the dataset.

Patient Selection

During the study period, a total of 54 patients who were fasting during Ramadan and presented to the emergency department with acute right upper quadrant pain or related biliary complaints were initially screened. Of these, 17 were excluded: 6 had no documented history of biliary system disease; 4 had missing laboratory data; 4 had no imaging performed during the visit; and 3 met exclusion criteria (1 with active malignancy, 1 on dialysis, 1 who was pregnant). The final analyzed sample consisted of 37 patients (Figure 1).

Inclusion and Exclusion Criteria

Patients with a history of gallstones, bile duct stones, and/or biliary sludge who had not been hospitalized in the last 6 months were selected. For the purposes of this study, "known biliary system disease" was defined as a previously documented diagnosis of cholelithiasis, choledocholithiasis, or biliary sludge, confirmed by prior imaging (USG, CT, or MRCP) or by surgical or procedural history [e.g., previous endoscopic retrograde cholangiopancreatography (ERCP)], and recorded in the patient's medical records or self-reported with supporting documentation. Patients on chronic medication, patients with diabetes or on dialysis, patients with missing blood parameters, patients without CT or USG images, patients with diseases causing non-biliary abdominal pain, patients under 18 years of age, pregnant women, and patients with a history of cancer were excluded from the study.

Statistical Analysis

Sample Size and Statistical Analysis

The G*Power 3.1.7.9 program was used to calculate the study sample size. The minimum sample size was determined to be 27, with the test power set at 0.80, the margin of error (alpha) at 0.05, and the effect size at 0.50. Data obtained in this study were analyzed using IBM SPSS 27.0. Descriptive statistics were calculated to evaluate the data. Normality was tested using the Kolmogorov-Smirnov test. Because the data were not normally distributed, non-parametric tests were used. Descriptive statistics for continuous variables were presented as median (Q1-Q3), minimum, and maximum; distributions of categorical variables were presented as frequencies and percentages (%). Mann-Whitney U and Kruskal-Wallis tests were used to evaluate the differences between groups because of the non-normal distribution of the data. For categorical comparisons, Fisher's exact test or Fisher–Freeman–Halton exact test was applied where expected cell counts were below 5, as the chi-square test may not be appropriate under these conditions. Results were evaluated at a 95% confidence level, and $p<0.05$ was considered statistically significant. Given the multiple comparisons performed across laboratory parameters and patient subgroups, all analyses should be considered exploratory. No formal correction for multiple testing was applied; findings are intended to generate hypotheses for future confirmatory studies rather than to establish definitive statistical inference.

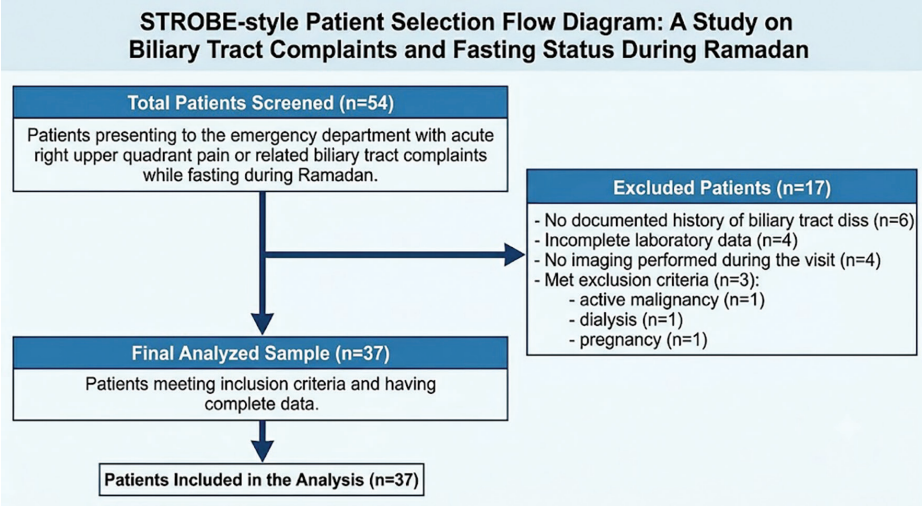


Figure 1. A strobe-style flowchart for patient selection

Results

The study included 37 patients. When descriptive statistics for continuous variables were calculated, the patients' median age was 62 years (Q1-Q3: 47-72), ranging from 28 to 91 years. The median WBC value was 8.80 (6-11.55). The median CRP level was 15 (6.50-58.00), with values ranging from a minimum of 1.00 to a maximum of 181. The median ALT was 30 (21.50-61.50), the median AST was 36 (23.50-67.00), and the median GGT was 63 (35-136), indicating a wide range of values (13-701). The median LDH was 250 (196-400), with a wide range of 150-716. The median value for total bilirubin was 2.40 (1.30-6.90), and for direct bilirubin was 1.30 (0.86-4.40).

The majority of the patients (62.2%) were female. USG/CT/MRCP reports indicated that 48.6% of patients had normal findings, 27% had a dilated CBD, 13.5% had gallstones, and 10.8% had biliary sludge. Regarding hospitalization status, 43.2% of the individuals were discharged, 27% were admitted to the internal medicine ward, 18.9% were admitted to the general surgery ward, and 10.8% were referred to other centers. All referrals were due to the unavailability of ERCP in our province (Figure 2).

A statistically significant difference was found between females and males in the AST variable (U=88.50; $p<0.05$). The results indicate that AST levels were significantly higher in women than in men. Furthermore, no statistically significant differences were found between the groups for age (p=0.280), WBC (p=0.790), CRP (p=0.125), ALT (p=0.742), GGT (p=0.065), LDH (p=0.091), total bilirubin (p=0.742), and direct bilirubin (p=0.975) (Figure 3).

No statistically significant differences were found in the biochemical parameters WBC, ALT, AST, and LDH according to the USG/CT/MRCP results. Since the data were not normally

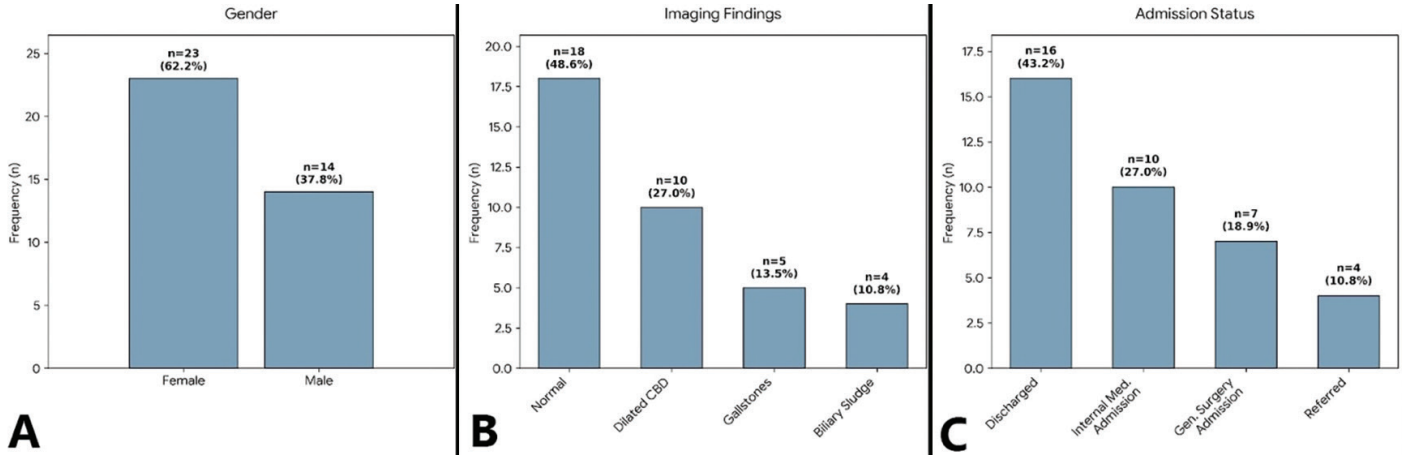


Figure 2. Frequency and percentage distributions of categorical variables
A) Gender, B) Imaging finding, C) Admission status

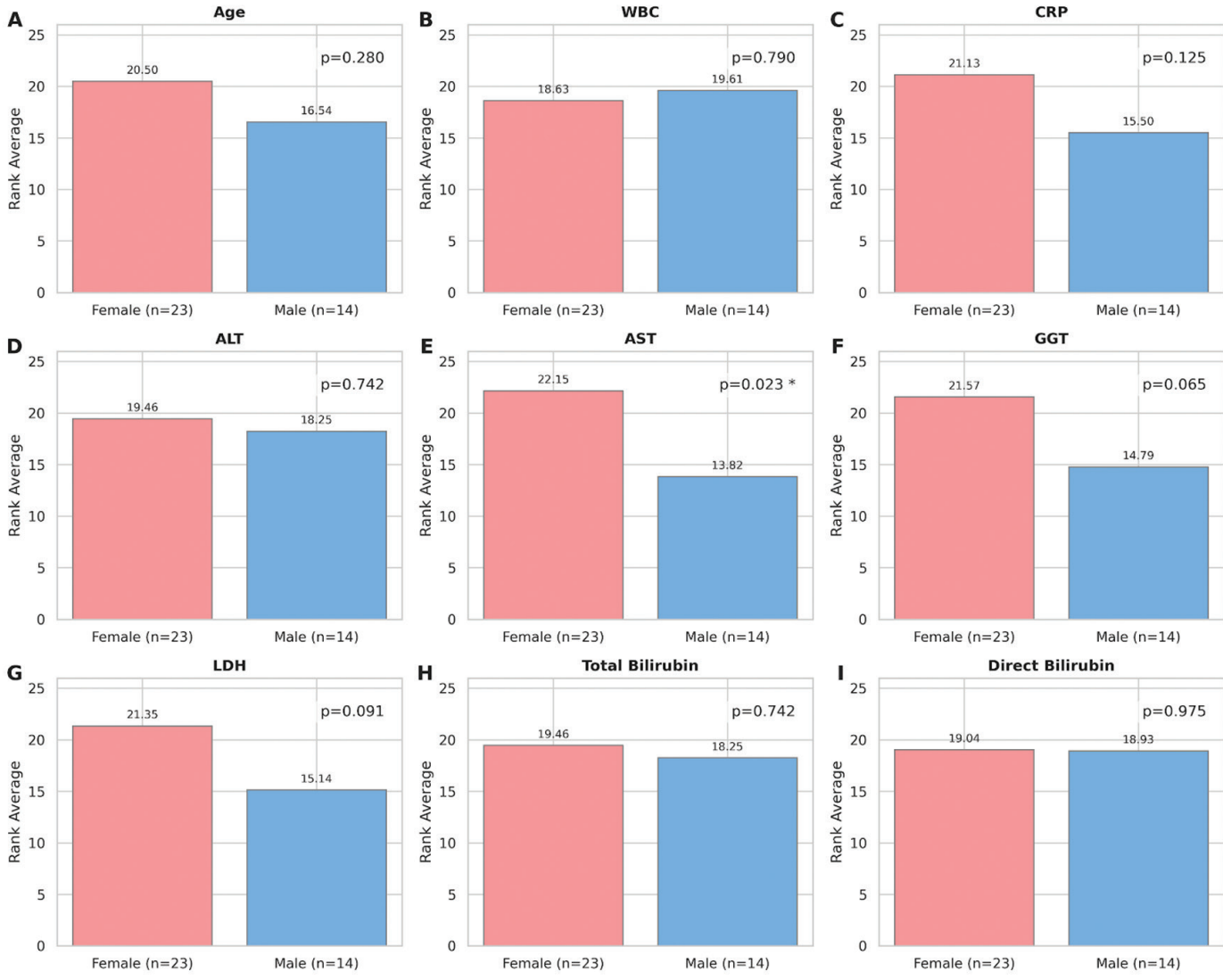


Figure 3. Mann-Whitney U test results for the comparison of biochemical parameters by gender

A) Age, B) WBC, C) CRP, D) ALT, E) AST, F) GGT, G) LDH, H) Total bilirubin, I) Direct bilirubin

The Mann-Whitney U test was applied to evaluate differences between gender groups in biochemical parameters that did not show a normal distribution. *p<0.05, **p<0.01, and ***p<0.001 were considered statistically significant

WBC: White blood cell count, CRP: C-reactive protein, ALT: Alanine aminotransferase, AST: Aspartate aminotransferase, GGT: Gamma-glutamyl transferase, LDH: Lactate dehydrogenase

distributed, the analysis was performed using the Kruskal-Wallis test. A statistically significant difference in CRP levels was found between the groups (H=8.543, p=0.036). Median CRP values were lower in the normal and dilated CBD groups and higher in the gallstone and biliary sludge groups. A significant difference in total bilirubin levels was found between the groups (H=22.139, p<0.001). Post hoc tests were conducted to determine which groups differed. Dunn-Bonferroni post hoc analysis indicated that these differences stemmed from the dilated CBD and biliary sludge groups (p=0.006). Significant differences were detected between the normal group and the biliary sludge (p=0.023), gallstone (p=0.005), and dilated CBD (p<0.001) groups. Total bilirubin levels in the dilated CBD and biliary sludge groups were higher than in the normal group. The gallstone group had lower levels than the dilated CBD and biliary sludge groups. Groups differed significantly in direct bilirubin levels (H=21.997, p<0.001). According to the Dunn-Bonferroni post hoc analysis, significant differences were observed between the normal imaging group and the biliary sludge (p=0.016), gallstone (p=0.007), and dilated CBD (p<0.001) groups. Direct bilirubin

levels were higher in the dilated CBD and biliary sludge groups than in the normal group. The gallstone group had a lower level than the dilated CBD and biliary sludge groups (Table 1).

A statistically significant difference between the groups was found for CRP, AST, LDH, total bilirubin, and direct bilirubin ($p<0.05$). Consequently, levels of CRP, AST, LDH, total bilirubin, and direct bilirubin differed by hospitalization status. Analysis using the Kruskal-Wallis test revealed a statistically significant difference between groups with respect to the relevant variables. Post hoc analysis was performed to determine which groups differed significantly. Regarding the CRP variable, a significant difference was found between the referred patients and patients admitted to the internal medicine service ($p=0.027$), and between the referred patients and those admitted to the general surgery service ($p=0.030$). Examination of CRP levels showed that median values were lower in the discharge and referral groups and higher in the internal medicine and general surgery admission groups. The difference in CRP by hospitalization status reflects a significant difference between these groups. For both total bilirubin and direct bilirubin, a significant difference ($p=0.010$) was found between those who were discharged and those admitted to internal medicine/general surgery services or referred. Specifically, AST levels were observed to be higher in the referral and internal medicine admission groups. LDH levels were notably higher in the internal medicine admission group. Total bilirubin and direct bilirubin levels were higher in the referral group (Table 2).

Examination of the Fisher–Freeman–Halton exact test results for hospitalization status according to USG, CT, and MRCP revealed a statistically significant difference between the groups ($p<0.001$). All referred patients had a dilated CBD, and MRCP revealed stones in the CBD, indicating the need for ERCP. Importantly, all referrals were necessitated solely by the unavailability of ERCP at our institution, and should not be interpreted solely as markers of clinical severity.

The Fisher–Freeman–Halton exact test results for gender by hospitalization status showed a statistically significant difference between the groups ($p=0.007$). It was observed that 43.5% of individuals admitted to the internal medicine service were women, whereas the discharge rate in male patients (64.3%) was higher than that in women. These results indicate that the need for hospitalization was significantly higher in female patients than in male patients (Table 3).

Discussion

Islam exempts patients with chronic illnesses from fasting. However, some individuals wish to fast, which may be associated with clinical deterioration in those who are medically medically unsuitable for fasting. Dietary habits change during the month of Ramadan (14). Generally, patients with diabetes mellitus and those with gastrointestinal diseases (ulcerative colitis, Crohn's disease, and peptic ulcer) are adversely affected by dietary changes, such as prolonged fasting. In the literature, beneficial effects of intermittent fasting, such as improving blood pressure, heart rate, cholesterol levels, and glycemic control, have also been reported (15). Additionally, it has been observed that fasting is associated with reduced incidence of diseases such

Table 1. Kruskal-Wallis H test results for the comparison of biochemical parameters according to USG/CT/MRCP results

Variables	Groups	Rank average	Median (min-max)	SD	H	p
CRP	Normal (n=18)	19.94	16.50 (5-181)	3	8,543	0,036*
	Dilated common bile duct (n=10)	11.85	8 (1-48)			
	Gallstones (n=5)	21.50	54 (2-155)			
	Bile sludge (n=4)	29.50	79 (20-176)			
Total bilirubin	Normal (n=18)	10.53	1.30 (0.80-2.80)	3	22,139	<0,001***
	Dilated common bile duct (n=10)	28.80	7.60 (2.40-12.60)			
	Gallstones (n=5)	25.80	4.80 (3.60-8.60)			
	Bile sludge (n=4)	24.13	8.40 (0.80-10.70)			
Direct bilirubin	Normal (n=18)	10.53	0.90 (0.40-1.40)	3	21,997	<0,001***
	Dilated common bile duct (n=10)	28.70	5.00 (1.30-9.20)			
	Gallstones (n=5)	25.40	3.30 (2.80-6.60)			
	Bile sludge (n=4)	24.88	6.20 (0.40-8.70)			

The Kruskal-Wallis test was applied to evaluate differences defined by USG/CT/MRCP results for biochemical parameters that were not normally distributed. *p<0.05, **p<0.01; ***p<0.001 were considered statistically significant
CRP: C-reactive protein, SD: Standard deviation, USG: Ultrasonography, CT: Computed tomography, MRCP: Magnetic resonance cholangiopancreatography

Table 2. Kruskal-Wallis H test results for the comparison of biochemical parameters according to hospitalization status

Variables	Groups	Rank average	Median (min-max)	SD	H	p
CRP	Discharged (n=16)	16.00	12.50 (5-170)	3	7,922	0,048*
	Referred (n=4)	9.63	7.50 (4-11)			
	Internal medicine admission (n=10)	23.80	47 (1-181)			
	General surgery hospitalization (n=7)	24.36	62 (2-176)			
AST	Discharged (n=16)	13.00	30.00 (10-55)	3	10,098	0,018*
	Referred (n=4)	20.13	160 (15-746)			
	Internal medicine admission (n=10)	26.50	76.50 (25-467)			
	General surgery hospitalization (n=7)	21.36	40 (26-95)			
LDH	Discharged (n=16)	12.09	196,00 (150-590)	3	12,471	0,006**
	Referred (n=4)	22.63	313 (200-550)			
	Internal medicine admission (n=10)	26.70	390,50 (228-520)			
	General surgery hospitalization (n=7)	21.71	258,00 (160-716)			
Total bilirubin	Discharged (n=16)	10.50	1.30 (0.80-4.20)	3	20,803	<0,001***
	Referred (n=4)	33.50	11.60 (5.41-12.60)			
	Internal medicine admission (n=10)	21.80	4.10 (1.00-10.60)			
	General surgery hospitalization (n=7)	26.14	5.80 (1.80-10.70)			
Direct bilirubin	Discharged (n=16)	10.66	0.90 (0.40-3.30)	3	19,988	<0,001***
	Referred (n=4)	33.25	8.05 (3.97-9.20)			
	Internal medicine admission (n=10)	21.80	3.05 (0.60-8.70)			
	General surgery hospitalization (n=7)	25.93	3.60 (1.12-7.90)			

The Kruskal-Wallis test was applied to evaluate differences between groups based on hospitalization status for biochemical parameters that were not normally distributed. *p<0.05, **p<0.01; ***p<0.001 were considered statistically significant
CRP: C-reactive protein, AST: Aspartate aminotransferase, LDH: Lactate dehydrogenase, SD: Standard deviation

Table 3. Chi-square test results for hospitalization status according to USG/CT/MRCP results

		Discharged	Referred	Internal medicine admission	General surgery hospitalization	Total	X^2	SD	p
USG/ CT/ MRCP	Normal	13 (81.3)	0 (0)	4 (40.0)	1 (14.3)	18 (48.6)	34.382	9	<0.001***
	Dilated common bile duct	2 (12.5)	4 (100.0)	4 (40.0)	0 (0)	10 (27.0)			
	Gallstones	0 (0)	0 (0)	1 (10.0)	4 (57.1)	5 (13.5)			
	Bile sludge	1 (6.3)	0 (0)	1 (10.0)	2 (28.6)	4 (10.8)			
Gender	Female	7 (30.4)	1 (4.3)	10 (43.5)	5 (21.7)		10.997	3	0.007**
	Male	9 (64.3)	3 (21.4)	0 (0)	2 (14.3)				

*p<0.05, **p<0.01; ***p<0.001 are statistically significant
USG: Ultrasonography, CT: Computed tomography, MRCP: Magnetic resonance cholangiopancreatography, SD: Standard deviation

as myocardial infarction, peptic ulcer, and heart failure (16,17). Understanding the associations between Ramadan fasting and clinical outcomes in patients with chronic conditions is crucial for both patients and physicians. Various fasting and nutritional patterns are used by individuals practicing intermittent fasting. The most popular is the "16-8" method; individuals do not consume any calories for 16 hours and use an 8-hour window for meals. Fasting during Ramadan is similar to this dietary method (2). Since diets of this nature restrict food intake for over 14 hours, glycogen stores are depleted first, followed by a decrease in cholesterol and triglyceride levels (18,19). Meo and Hassan (20) reported that Ramadan fasting reduces risk factors for proinflammatory and atherosclerotic diseases by causing a decrease in inflammation and proinflammatory cytokines such as IL-1b, IL-6, and TNFa. Physiologically, restricting food intake allows the digestive system to rest by restructuring the

gut microbiome. It also stimulates autophagy, supporting the elimination of impaired cellular components and the renewal of cells. When the migrating motor complex, responsible for intestinal cleansing, is activated, smooth muscle contractions occur, gastric acid is secreted, and particles pass into the small intestine. Pancreatic secretions and bile from the gallbladder are released to neutralize gastric acid. As bile fluid passes through the gastrointestinal system, it kills bacteria and prevents them from adhering to the intestinal wall (21). Consequently, studies investigating the association between prolonged fasting and enzymes released from the liver and bile have been conducted (11,12). Rapid consumption of high-calorie foods at iftar (the breaking of the fast) following prolonged hunger may be associated with symptomatic exacerbation in patients with biliary disorders. Skipping meals or prolonged fasting has been associated with reduced gallbladder emptying, thereby facilitating stone formation (22). During fasting, the gallbladder becomes hypotonic; bile volume and concentration increase, and its (bile) fluidity decreases. At iftar, the rapid consumption of foods that are usually high in fat and carbohydrates strongly stimulates the release of cholecystokinin. This may cause the gallbladder, filled with concentrated bile, to contract against a possible obstruction (stone or sludge), potentially contributing to the emergence of symptoms (21). It must be emphasized, however, that because no control group was included in this study, a direct causal relationship between Ramadan fasting and biliary symptom exacerbation cannot be established; the observed associations must be interpreted as hypothesis-generating. Consistent with the literature, a study conducted in Egypt reported fluctuations in liver enzymes in fasting patients but no significant change in bilirubin levels (22). In our study, bilirubin levels were significantly higher, particularly in the dilated CBD and biliary sludge groups. Importantly, this finding most likely reflects the greater anatomical severity and obstructive nature of the disease in these patients rather than a direct effect of fasting per se. Elevated bilirubin in the context of CBD dilatation and biliary sludge is a recognized marker of biliary obstruction severity; the emergency department setting inherently selects for patients who are more acutely unwell.

In the initial hours of fasting, there may be an increase in values indicating the risk of gallstones; as the body adapts to fat burning, the sudden release of this fat into the bloodstream can temporarily increase cholesterol levels, affecting bile composition (23). Hunger and dehydration have been shown to be associated with acalculous cholecystitis (12). In our study, there were no significant differences between the groups for ALT and AST values. There were no changes in AST, ALT, LDH, or GGT values in patients with calculous cholecystitis. The AST value was significantly higher in women than in men ($p<0.05$). CRP levels were significantly higher in patients with gallstones and biliary sludge. Total and direct bilirubin values were higher in the dilated CBD and biliary sludge groups than in the normal group. These laboratory associations most likely reflect the inherent severity of the underlying biliary pathology rather than a direct consequence of fasting, given the absence of pre-fasting or non-fasting comparison data. The female gender dominance (62.2%) in our study is noteworthy, and this finding aligns with the literature stating that gallstones are more prevalent in women (24). Hospitalization rates among female patients were significantly higher than among male patients ($p=0.007$). This association in female patients may reflect a more severe symptomatic course or a higher risk of complications (cholecystitis, pancreatitis, and other complications), though firm conclusions regarding causality require appropriately controlled prospective studies. It is reported that cholecystitis and CBD stone patients constitute approximately 15% of general surgery admissions (25). In our study, a similar proportion of patients (18.9%) were admitted to the general surgery service. On the other hand, it is reported that approximately 20% of patients with biliary sludge receive follow-up in internal medicine services because they require conservative treatment and ERCP (26). In our study, the admission rate to the internal medicine service was 27%. Median CRP values were lower in discharged and referred patients but higher in patients admitted to hospital wards. Furthermore, AST levels were higher in referred patients and in those admitted to the internal medicine service. Total bilirubin and direct bilirubin levels were also higher in the referral group. Collectively, these findings suggest that elevated CRP and bilirubin may serve as markers of disease severity in this clinical context, rather than being directly attributable to the fasting state.

Study Limitations

The primary limitations of our study are its single-center design and its relatively small sample size ($n=37$), which together constrain statistical power, particularly in subgroup analyses, and limit the generalizability of the findings. The most critical methodological limitation is the absence of a non-fasting or pre-Ramadan control group. Without a comparator cohort, it is not possible to determine whether the observed laboratory alterations and hospitalization rates are attributable to fasting itself or to the expected clinical course of biliary disease in this population. This study should therefore be regarded as descriptive and hypothesis-generating rather than inferential. Important confounding variables, including fasting duration, degree of dehydration, dietary content during iftar and suhoor, body mass index (BMI), and concurrent medications, were not systematically recorded or adjusted for. The inability to standardize nutritional habits during Ramadan further limits the interpretability of

associations between dietary patterns and clinical outcomes. Additionally, the exclusion of post-hospitalization follow-up data and post-discharge recurrence rates prevents the assessment of long-term effects. The study's other limitations include the absence of data on several clinically significant variables, such as patients' fever at admission, Murphy's sign, time elapsed from iftar or sahur to symptom onset, BMI, detailed comorbidity profiles, prior cholecystectomy or ERCP, final diagnosis at discharge, results of ERCP procedures for referred patients, length of hospital stay, and post-discharge recurrence rates.

Conclusion

This descriptive observational study suggests that Ramadan fasting may be associated with laboratory and clinical changes in patients with diseases of the biliary system, though no causal relationship can be established given the study design. In patients with biliary pathology presenting to the emergency department, the need for hospitalization and advanced investigation (ERCP) was high, particularly among those with elevated CRP and bilirubin levels. Female gender appears to be associated with higher hospitalization rates in this patient group. Clinically, patients with known biliary system diseases (e.g., gallstones or sludge) should be evaluated for individual risk prior to Ramadan; symptomatic patients should receive appropriate counselling regarding fasting; and dietary modifications should be implemented where appropriate. These findings underscore the need for well-designed, multicenter, controlled prospective studies with larger sample sizes to more rigorously assess the relationship between Ramadan fasting and biliary disease outcomes.

Ethics

Ethics Committee Approval: The study was initiated following the approval of the Kafkas Medical Faculty Ethics Committee (decision number: 02, date: 27.02.2024).

Informed Consent: Signatures were obtained from all patients for the informed consent form.

Footnotes

Authorship Contributions

Surgical and Medical Practices: L.Ş., T.G., Concept: L.Ş., Design: M.A.A., Data Collection or Processing: L.Ş., Analysis or Interpretation: M.A.A., Literature Search: T.G., Writing: L.Ş.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Köktürk SN, Yardımcı H. Aralıklı oruç ve bazı hastalıklar üzerine etkileri. Turkiye Klinikleri J Health Sci. 2021;6:949-57.
2. Tibi S, Ahmed S, Nizam Y, Aldoghmi M, Moosa A, Bourenane K, et al. Implications of Ramadan fasting in the setting of gastrointestinal disorders. Cureus. 2023;15:e36972.
3. Onur A, Caferoğlu Z, Hatipoğlu N. Ramazan ayında oruç tutan diyabetli bireylerde diyabetin yönetimi. Turk J Diab Obes. 2021;5:59-67.
4. Hosseini SV, Torabijahromi M, Mosallaei M, Sabet B, Pourahmad S. The effect of season and Ramadan fasting on the onset of acute cholecystitis. Saudi Med J. 2006;27:503-6.
5. Emami-Naini A, Roomizadeh P, Baradaran A, Abedini A, Abtahi M. Ramadan fasting and patients with renal diseases: a mini review of the literature. J Res Med Sci. 2013;18:711-6.
6. Göçmen E, Koç M, Tez M, Yoldaş O, Bilgin A, Keşkek M. Effect of Ramadan on surgical emergencies. Ann Emerg Med. 2004;44:283-5.
7. Persynaki A, Karras S, Pichard C. Unraveling the metabolic health benefits of fasting related to religious beliefs: a narrative review. Nutrition. 2017;35:14-20.
8. Moore J, Fung J. The complete guide to fasting: heal your body through intermittent, alternate-day, and extended fasting. Simon and Schuster. 2016.
9. Boobes Y, Afandi B, AlKindi F, Tarakji A, Al Ghamdi SM, Alrukhaimi M, et al. Consensus recommendations on fasting during Ramadan for patients with kidney disease: review of available evidence and a call for action (RaK Initiative). BMC Nephrol. 2024;25:84
10. Abdolreza N, Omalbanin A, Mahdieh TS, Mohammad Ali MR, Reza MS, Maryam S, et al. Comparison of the number of patients admitted with renal colic during various stages of peri-Ramadan month. Saudi J Kidney Dis Transpl. 2011;22:1199-202.
11. Elfert AA, AbouSaif SA, Abdel Kader NA, AbdelAal E, Elfert AY, Abdel Moez AT, et al. A multicenter pilot study of the effects of Ramadan fasting on patients with liver cirrhosis. Tanta Med Sci J. 2011;6:33-6.
12. Kuroi Y, Imazato D, Yamazaki K, Kasuya H. Acute cholecystitis in patients with stroke. Neurol India. 2019;67:439-41.
13. Samuel I. Negative impact of prolonged fasting on gallbladder ejection fraction measurement with cholescintigraphy in chronic acalculous symptomatic hyPERkinetic (CASPER) gallbladders. Medical Research Archives. 2024;12.
14. Şimşek P, Yadigaroğlu M, Katipoğlu B, Şahin A, Topbaş M, Gündüz A. Effect of Ramadan on emergency department attendances: a retrospective study. Eurasian J Emerg Med. 2021;20:219-23.
15. Li Z, Heber D. Intermittent Fasting. JAMA. 2021;326:1338.
16. Jalali F, Hajian T, Laki KO. Day of week, monthly and seasonal variation of acute myocardial infarction. Acta Medica Iranica. 2002;4:230-5.
17. Akhan G, Kutluan S, Koyuncuoglu HR. Is there any change of stroke incidence during Ramadan ?. Acta Neurol Scand. 2000;101:259-61.
18. Memel ZN, Wang J, Corey KE. Intermittent fasting as a treatment for nonalcoholic fatty liver disease: what ıs the evidence? Clin Liver Dis (Hoboken). 2022;19:101-5.
19. Nematy M, Alinezhad-Namaghi M, Rashed MM, Mozhdehifard M, Sajjadi SS, Akhlaghi S, et al. Effects of Ramadan fasting on cardiovascular risk factors: a prospective observational study. Nutr J. 2012;11:69.
20. Meo SA, Hassan A. Physiological changes during fasting in Ramadan. J Pak Med Assoc. 2015;65:S6-14.
21. Deloose E, Janssen P, Depoortere I, Tack J. The migrating motor complex: control mechanisms and its role in health and disease. Nat Rev Gastroenterol Hepatol. 2012;9:271-85.

22. Nasiri J, Kheiri S, Khoshdel A, Boroujeni AJ. Effect of Ramadan fast on liver function tests. Iran J Med Sci. 2016;41:459-60.
23. Zhang H, Xu C, Zhu X, Zhang J, Yin J, Yao N, et al. Associations between temporal eating patterns and energy distribution patterns with gallstones: a cross-sectional study based on NHANES 2017-2018. BMC Public Health. 2024;24:2994.
24. Lazarchuk I, Barzak B, Wozniak S, Mielczarek A, Lazarchuk V. Cholelithiasis–a particular threat to women. A review of risk factors. Medical Journal of Cell Biology. 2023;11:20-7.
25. Yew S, Cheong CM, Glen P. 43 Burden of biliary pathologies on emergency surgical admissions. British Journal of Surgery. 2025;112.
26. Loozen CS, van Ramshorst B, van Santvoort HC, Boerma D. Early cholecystectomy for acute cholecystitis in the elderly population: a systematic review and meta-analysis. Dig Surg. 2017;34:371-9.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.03764
Eurasian J Emerg Med. 2026;25: 360-5

Structural Factors Associated with Provincial Variation in False Emergency Medical Service Calls in Türkiye: An Ecological Analysis

Çağdaş Derdiyok

Lüleeburgaz State Hospital, Clinic of Emergency Medicine, Kırklareli, Türkiye

Abstract

Aim: To evaluate provincial variation in false emergency medical service (EMS) call rates across Türkiye and to examine the association between education indicators and these rates after accounting for EMS accessibility and primary care availability.

Materials and Methods: This ecological study analyzed province-level data for 2024 from all 81 Turkish provinces using official datasets from the Ministry of Health General Directorate of Emergency Health Services, the Turkish Statistical Institute, and the Health Statistics Yearbook 2024. The dependent variable was the provincially reported false alarm rate (%), defined as a composite administrative measure including non-emergency conditions manageable in primary care, hoax or malicious calls, and refused transport after ambulance arrival. These three components likely reflect distinct underlying mechanisms: non-emergency calls may represent barriers to primary care access, hoax calls may reflect behavioral and social factors, and refused transport may indicate that the caller has reassessed urgency or is seeking alternative care. Because these components could not be analyzed separately in the available aggregated data, the composite outcome was interpreted cautiously as a system-level operational indicator, rather than as a homogeneous clinical or behavioral construct. Inter-provincial differences in triage protocols and documentation practices may further affect the consistency of classification. Education indicators, EMS accessibility measures, and primary care availability were examined. Pearson correlation analyses, independent-samples t-tests, and multiple linear regression analyses with robust standard errors were performed.

Results: The mean false alarm rate was 8.75%±2.86% (range 4.2%-14.7%). Metropolitan provinces had significantly higher rates than non-metropolitan provinces (10.12%±3.45% vs. 7.88%±1.89%, $p<0.001$). The university education rate showed a weak, non-significant correlation with the false alarm rate ($r=0.13$, $p=0.258$). Population per family physician unit showed a moderate positive correlation with the false alarm rate ($r=0.545$, $p<0.001$). In the multiple linear regression model ($R^2=0.378$, $p<0.001$), the university education rate was not independently associated with false alarm rates ($\beta=-0.062$, $p=0.434$), although the association's direction was negative (suggesting lower false alarm rates with higher education). Population served per EMS station ($\beta=0.134$, $p=0.006$) and population per family physician unit ($\beta=7.317$, $p<0.001$) showed independent, statistically significant associations. The number of EMS stations was not independently associated ($\beta=0.006$, $p=0.477$).

Conclusion: In this nationwide ecological analysis, provincial education levels were not independently associated with rates of false EMS calls after adjustment for EMS accessibility and primary care availability. Both EMS accessibility and primary care availability showed significant, independent associations with false alarm rates, suggesting that structural factors related to access to healthcare services are more strongly associated with provincial variation than educational indicators. The findings are consistent with the possibility that limited access to primary care may contribute to increased non-urgent EMS utilization at the population level.

Keywords: Emergency medical services, false EMS calls, EMS utilization, health service accessibility, primary care, ecological study, Türkiye

Corresponding Author: Çağdaş Derdiyok MD, Lüleeburgaz State Hospital, Clinic of Emergency Medicine, Kırklareli, Türkiye
E-mail: drcagdasderdiyok@gmail.com **ORCID ID:** orcid.org/0000-0002-0651-4649

Cite this article as: Derdiyok Ç. Structural factors associated with provincial variation in false emergency medical service calls in Türkiye: an ecological analysis. Eurasian J Emerg Med. 2026;25: 360-5.

Received: 03.04.2026
Accepted: 08.06.2026
Published: 26.06.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Introduction

Emergency medical services (EMS) are designed to provide rapid response to life-threatening conditions. However, non-urgent or inappropriate EMS utilization remains a persistent challenge worldwide and may contribute to increased workload, inefficient resource allocation, and potential delays for critically ill patients (1,2). Understanding the structural determinants of system-level EMS utilization patterns has therefore become an important focus of health services research.

Reported rates of non-urgent EMS use vary substantially across countries and healthcare systems (3,4). These variations are associated not only with individual patient behavior but also with structural characteristics such as service accessibility, dispatch organization, urbanization level, and primary care availability (5-9). Previous research has consistently reported that limited access to primary care services is correlated with higher non-urgent EMS utilization, as patients may use emergency services as an alternative access pathway when primary care is unavailable or inaccessible (2,4).

In Türkiye, EMS services are coordinated through a nationwide centralized dispatch structure via the 112 Emergency Call Centers, providing a standardized organizational framework suitable for evaluating inter-provincial variation using national administrative EMS datasets (10). Primary care services are delivered through the Family Medicine system, with each family physician unit serving a defined population (11). This ecological study therefore aimed to evaluate provincial variation in false EMS call rates across all 81 provinces of Türkiye and to examine whether education indicators were independently associated with these rates after accounting for EMS accessibility measures and primary care availability. As with all ecological analyses, interpretation of province-level associations should consider the potential for ecological fallacy (12).

Materials and Methods

This study used publicly available province-level aggregated administrative data and did not include individual patient-level information. Therefore, ethics committee approval and individual informed consent were not required.

Study Design and Data Sources

This ecological study analyzed province-level aggregate data from all 81 provinces of Türkiye using publicly available 2024 national datasets. Data on provincially reported false EMS call rates and the number of emergency health stations were obtained from the Ministry of Health General Directorate of Emergency Health Services (10). Education attainment data were obtained from the Turkish Statistical Institute National Education Statistics Database (13). Primary care availability data, including family physician unit numbers and population per family physician unit, were obtained from the Health Statistics Yearbook 2024, Table 7.20 (11).

Statistical Analysis

Statistical analyses were performed using Python (version 3.12; pandas, SciPy, and statsmodels libraries). Descriptive statistics are presented as mean ± standard deviation (SD). Bivariate associations were examined using Pearson correlation coefficients. Multiple linear regression with Huber-White robust standard errors was used to assess independent associations, with the dependent variable treated as a continuous proportion bounded between 0 and 1. Although fractional logit and beta regression are alternative modeling approaches for proportional outcomes, linear regression with robust standard errors provides consistent estimates of marginal effects and has been widely used in ecological and health services research involving bounded proportional outcomes, and was selected for direct comparability with prior ecological EMS utilization studies and ease of coefficient interpretation. Sensitivity analyses using fractional logit regression yielded qualitatively similar results (Appendix 1), and the primary findings were therefore considered robust to the choice of modeling framework. For the fractional logit sensitivity analysis, all percentage values were divided by 100 and modelled as fractional values on the 0-1 scale, as required by the bounded mean assumption of the fractional response model (14). Multicollinearity was evaluated using variance inflation factors (VIF).

Results

Provincial Characteristics

Table 1 presents descriptive statistics for the 81 provinces included in the analysis. The mean false alarm rate was 8.75% (SD=2.86%, range 4.2%-14.7%). The mean population per family physician unit was 2.89 thousand (SD=0.17, range 2.54-3.25), with a national average of 2.97 thousand according to the Health Statistics Yearbook 2024 (11).

Metropolitan provinces (n=31) had significantly higher false alarm rates than non-metropolitan provinces (n=50) (10.12%±3.45% vs. 7.88%±1.89%, $p<0.001$).

Bivariate Associations

Table 2 presents the correlation matrix for all study variables. Population per family physician unit showed a moderate positive correlation with false alarm rate ($r=0.545$, $p<0.001$), indicating that provinces with lower primary care accessibility (higher

population per family physician unit) experienced higher false EMS call rates. Population served per EMS station also showed a moderate positive correlation with false alarm rate (r=0.459, p<0.001). The university education rate showed a weak, non-significant correlation with the false alarm rate (r=0.13, p=0.258) (Figure 1).

Multiple Regression Analysis

Table 3 presents the results of the multiple linear regression analysis. After adjustment for EMS accessibility measures and primary care availability, university education rate was not independently associated with false alarm rates [β=-0.062, standard error (SE)=0.079, p=0.434], although the direction of the association was negative (suggesting lower false alarm rates with higher education). In contrast, both structural accessibility indicators showed statistically significant independent associations: population served per EMS station (β=0.134, SE=0.049, p=0.006) and population per family physician unit (β=7.317, SE=1.797, p<0.001). Number of EMS stations was not independently associated with false alarm rates (β=0.006, SE=0.008, p=0.477). The overall model explained 37.8% of the variance in false alarm rates (R^2=0.378, F=15.12, p<0.001). VIF were all below 2.0, indicating no substantial multicollinearity. The distribution of false alarm rates across Turkish provinces is shown in Figure 2.

Discussion

After adjustment for EMS accessibility indicators and primary care availability, provincial university education rate was not independently associated with false EMS call rates. In contrast, both EMS accessibility-related structural indicators and primary care availability remained significantly associated with inter-provincial variation. These findings suggest that crude associations between education indicators and false EMS calls

Table 1. Provincial characteristics and key indicators (n=81)

Variable	Mean ± SD	Min	Max
False alarm rate (%)	8.75±2.86	4.20	14.66
University and higher education rate (%)	14.26±3.69	8.08	23.40
Illiteracy rate (%)	2.87±1.15	0.99	4.96
Upper secondary education rate (%)	17.37±2.25	12.64	22.31
Population served per EMS station (1000s)	17.73±6.64	6.51	36.16
Number of EMS stations	50.79±39.00	8.00	149.44
Population per family physician unit (1000s)	2.89±0.17	2.54	3.25

EMS: Emergency medical service, SD: Standard deviation

Table 2. Correlation matrix of education indicators, EMS accessibility, primary care availability, and false alarm rates

Variable	1	2	3	4	5	6	7
1. False alarm rate	1.00						
2. University education rate	0.13	1.00					
3. Illiteracy rate	-0.05	-0.80***	1.00				
4. Upper secondary rate	0.10	0.82***	-0.69***	1.00			
5. Pop per EMS station	0.46***	0.15	0.00	-0.01	1.00		
6. Number of EMS stations	0.09	0.54***	-0.45***	0.40***	-0.06	1.00	
7. Pop per family physician	0.55***	0.27*	-0.19	0.15	0.38***	0.18	1.00

*p<0.05, ***p<0.001, EMS: Emergency medical service

Table 3. Multiple linear regression analysis of factors associated with false alarm rates

Variable	β	SE	t	p	95% CI
Constant	-14.158	4.634	-3.055	0.002	-23.240 to -5.075
University education rate (%)	-0.062	0.079	-0.782	0.434	-0.217 to 0.093
Population served per station (1000s)	0.134	0.049	2.763	0.006	0.039 to 0.229
Number of EMS stations	0.006	0.008	0.711	0.477	-0.010 to 0.022
Population per family physician unit (1000s)	7.317	1.797	4.072	<0.001	3.795 to 10.839

R^2=0.378, F=15.12, p<0.001, EMS: Emergency medical service, SE: Standard error, CI: Confidence interval

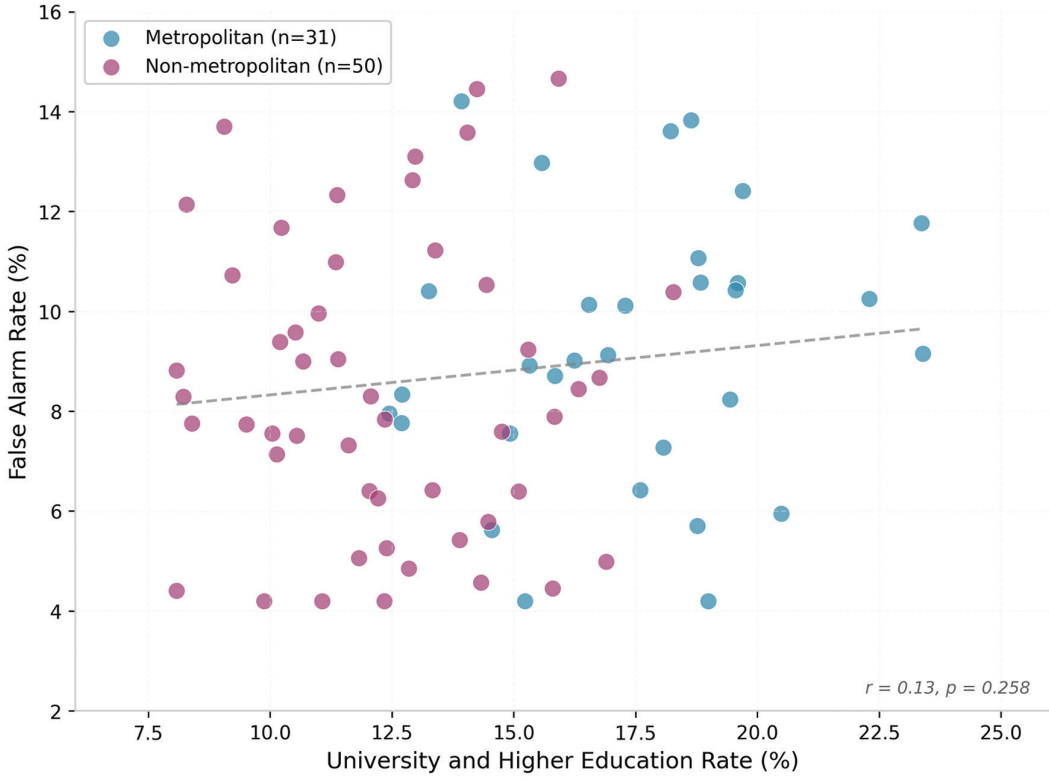


Figure 1. Relationship between university education rate and false alarm rates by province type

Metropolitan provinces (blue, n=31) had higher education rates but showed similar false alarm patterns compared to non-metropolitan provinces (purple, n=50). The correlation was not statistically significant (r=0.13, p=0.258)

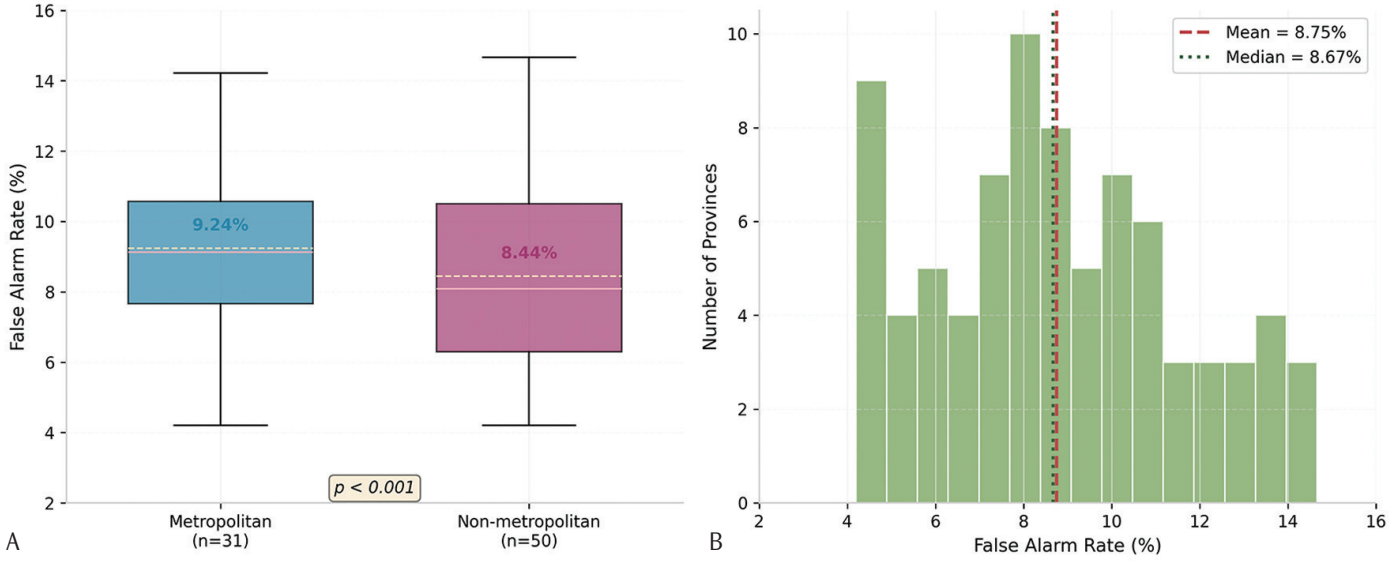


Figure 2. Distribution of false alarm rates across Turkish provinces

(A) Boxplot comparing metropolitan and non-metropolitan provinces showing significantly higher rates in metropolitan areas (p<0.001) Metropolitan median: 9.24%, non-metropolitan median: 8.44% (B) Histogram of false alarm rate distribution across all 81 provinces with overall mean (8.75%) and median (8.67%) indicated

are more likely to reflect differences in service accessibility and healthcare infrastructure than differences in education level itself.

The finding that population per family physician unit was independently associated with false EMS call rates represents an important contribution of this analysis. This finding is compatible with prior literature suggesting that limited access to primary care services may be associated with greater non-urgent EMS utilization. Previous studies have consistently reported that inadequate primary care availability is correlated with inappropriate emergency service use, as patients may resort to EMS when timely primary care access is unavailable (2,4,15). Our results extend these findings to the Turkish context, observing that provinces with better primary care accessibility tend to have lower false EMS call rates.

The magnitude of the association is substantial at the provincial level: each 1000-person increase in population per family physician unit was associated with approximately 7.3 percentage points higher false alarm rate, after controlling for EMS accessibility and education. This finding suggests that provinces with the lowest primary care accessibility (3.25 thousand population per unit) had false alarm rates approximately 5.2 percentage points higher than provinces with the best accessibility (2.54 thousand population per unit), based on model predictions.

The non-significant association between number of EMS stations and false alarm rates, after controlling for population served per station and primary care availability, may reflect the endogenous nature of EMS station allocation based on historical demand patterns. This finding is consistent with the reverse causality concern that EMS stations are typically allocated according to historical service demand rather than being an exogenous determinant of utilization patterns.

The inclusion of primary care accessibility indicators represents an important strength of this analysis compared with previous ecological studies of EMS utilization. By adjusting for both EMS and primary care structural characteristics, our model provides a more comprehensive assessment of the structural determinants of provincial variation in false EMS calls. However, the cross-sectional design precludes determination of whether improvements in primary care accessibility would lead to reductions in false EMS call rates over time.

From a policy perspective, the findings suggest that interventions focusing solely on general education level are unlikely to be associated with substantially lower false EMS call rates at the population level. Instead, system-level demand management strategies—such as improved public guidance on appropriate EMS use, telephone triage systems, community paramedicine programs, and strengthened after-hours primary care access—may provide more effective approaches for optimizing EMS utilization patterns. The significant association between primary care availability and false EMS calls suggests that provinces with lower population-to-family-physician ratios tended to demonstrate lower false EMS call rates.

Study Limitations

This study has several important limitations. First, the ecological design limits interpretation to province-level associations and does not allow individual-level inference. The ecological fallacy risk (14) is particularly relevant for the education-related findings: provincial-level education rates may not reflect the educational characteristics of individual EMS callers, and the absence of an association at the aggregate level does not preclude individual-level associations between health literacy and appropriate EMS use. Second, the dependent variable represents a composite administrative indicator combining non-emergency calls, hoax or malicious calls, and refused transport after ambulance arrival. These components likely reflect different underlying behavioral and structural mechanisms and could not be analyzed separately due to data aggregation. This aggregation may attenuate or mask associations that would be detectable if the components were analyzed separately; for example, if primary care accessibility is strongly associated with non-emergency calls but only weakly associated with hoax calls, the composite measure would dilute the true magnitude of that relationship. Third, the cross-sectional design precludes any causal inference; all reported associations are observational and may be subject to unmeasured confounding. Fourth, several potentially important confounders could not be included due to data availability, including provincial socioeconomic status indicators (unemployment rates, income levels), urbanization gradients beyond the metropolitan versus non-metropolitan dichotomy, regional variations in EMS triage protocols and documentation practices, cultural differences in healthcare-seeking behavior, and after-hours primary care availability. Fifth, although population per family physician unit was included as a proxy for primary care accessibility, this measure does not capture geographic accessibility within provinces, quality of primary care services, or informal care networks. Residual confounding related to unmet primary care demand and unobserved provincial characteristics therefore cannot be fully excluded (2,4). Finally, inter-provincial differences in how false calls are classified and reported may introduce measurement heterogeneity that could affect the observed associations.

Conclusion

In this nationwide ecological analysis including all 81 provinces of Türkiye, provincial university education rate was not independently associated with false EMS call rates after adjustment for EMS accessibility indicators and primary care availability. Both EMS accessibility measures and primary care availability showed significant independent associations with inter-provincial variation in false EMS call patterns, with population per family physician unit demonstrating the strongest association among the structural factors examined.

These findings suggest that system-level demand management strategies—such as improving guidance on appropriate EMS use, strengthening primary care accessibility, and implementing alternative triage pathways—may represent more effective approaches than interventions focusing solely on general education level. Region-specific planning approaches may be particularly important given the observed differences between metropolitan and non-metropolitan provinces, and provinces with lower population-to-family-physician ratios tended to demonstrate lower false EMS call rates.

Ethics

Ethics Committee Approval: This study used publicly available province-level aggregated administrative data and did not include individual patient-level information. Therefore, ethics committee approval were not required.

Informed Consent: Patient consent was not required, as this study used publicly available aggregated administrative data and did not include individual patient-level data.

Footnotes

Financial Disclosure: The author declared that this study received no financial support.

Appendix: https://d2v96fxpocvxx.cloudfront.net/8a9ff4da-541a-42fa-9980-1a9a3ab6d6c5/content-images/3d1a5056-178d-402b-8660-e7dd238196f7.pdf

References

1. Snooks H, Wrigley H, George S, Thomas E, Smith H, Glasper A. Appropriateness of use of emergency ambulances. J Accid Emerg Med. 1998;15:212-5.
2. Booker MJ, Shaw AR, Purdy S. Why do patients with 'primary care sensitive' problems access ambulance services? A systematic mapping review of the literature. BMJ Open. 2015;5:e007726.
3. Shah MN, Glushak C, Karrison TG, Mulliken R, Walter J, Friedmann PD, et al. Predictors of emergency medical services utilization by elders. Acad Emerg Med. 2003;10:52-8.
4. Afilalo J, Marinovich A, Afilalo M, Colacone A, Léger R, Unger B, et al. Nonurgent emergency department patient characteristics and barriers to primary care. Acad Emerg Med. 2004;11:1302-10.
5. Ruger JP, Richter CJ, Lewis LM. Clinical and economic factors associated with ambulance use to the emergency department. Acad Emerg Med. 2006;13:879-85.
6. Morgans A, Burgess S. What is the demand for ambulance services? Emerg Med Australas. 2012;24:507-13.
7. McConnel CE, Wilson RW. The demand for prehospital emergency services in an aging society. Health Serv Res. 1998;33:591-609.
8. Cutler DM, Lleras-Muney A. Understanding differences in health behaviors by education. J Health Econ. 2010;29:1-28.
9. van der Heide I, Wang J, Droomers M, Spreeuwenberg P, Rademakers J, Uiters E. The relationship between health literacy and health behaviors. J Health Commun. 2013;18 Suppl 1:172-84.
10. Republic of Türkiye Ministry of Health. Emergency Health Services Annual Report 2024. Ankara: Ministry of Health; 2024. (Available from: https://shgm.saglik.gov.tr/)
11. Republic of Türkiye Ministry of Health. Health Statistics Yearbook 2024. Ankara: Ministry of Health; 2024. (Available from: https://veriportali.tuik.gov.tr/en/press/53937)
12. Greenland S, Robins J. Invited commentary: ecologic studies--biases, misconceptions, and counterexamples. Am J Epidemiol. 1994;139:747-60.
13. Turkish Statistical Institute (TÜİK). National Education Statistics Database 2024. Ankara: TÜİK; 2024. Available from: https://veriportali.tuik.gov.tr/en/press/53937
14. Papke LE, Wooldridge JM. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. J Appl Econometrics. 1996;11:619-32.
15. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav. 1995;36:1-10.

EURASIAN JOURNAL OF EMERGENCY MEDICINE

DOI: 10.4274/eajem.galenos.2026.e001

Erratum

Eurasian J Emerg Med. 2026;25: 366

DOI: 10.4274/eajem.galenos.2026.02525

In the first page of the published article, the affiliation number of the author Durdane Yılmaz Güven was incorrectly indicated as 1. The affiliation information has been corrected by assigning affiliation number 2 to this author. The affiliation number of Durdane Yılmaz Güven has been corrected from 1 to 2.

Incorrect institution information of Durdane Yılmaz Güven (institution number 1).

Seda Öztürk, Durdane Yılmaz Güven

Karabük University, Graduate Education Institute, Nursing Science, Karabük, Türkiye

Corrected institution information of Durdane Yılmaz Güven (institution number 2).

Affiliation number 1 has been added for Seda Öztürk, and affiliation number 2 has been added for Durdane Yılmaz

Seda Öztürk[1], Durdane Yılmaz Güven[2]

[1]Karabük University, Graduate Education Institute, Nursing Science, Karabük, Türkiye

[2]Karabük University, Faculty of Health Sciences, Department of Nursing, Karabük, Türkiye

Incorrect corresponding author information for Durdane Yılmaz Güven.

Durdane Yılmaz Güven MD, Karabük University, Graduate Education Institute, Nursing Science, Karabük, Türkiye

Correct corresponding author information for Durdane Yılmaz Güven.

Durdane Yılmaz Güven MD, Karabük University, Faculty of Health Sciences, Department of Nursing, Karabük, Türkiye

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.06787
Eurasian J Emerg Med. 2026;25: 367-74

Mortality Prediction in Geriatric Hip Fracture Patients in the Emergency Department: A Comparison of HALP, mHALP, and mREMS Scores

Abdülaziz Doğan[1], Ali Ünal[2]

[1]University of Health Sciences Türkiye, Konya City Hospital, Clinic of Emergency Medicine, Konya, Türkiye
[2]Uşak University Faculty of Medicine, Department of Emergency Medicine, Uşak, Türkiye

Abstract

Aim: To compare the performance of hemoglobin, albumin, lymphocyte, and platelet (HALP), modified HALP (mHALP), and modified Rapid Emergency Medicine score (mREMS) in predicting 30-day mortality in geriatric patients presenting to the emergency department (ED) with hip fractures.

Materials and Methods: This prospective, observational cohort study included 63 geriatric patients with hip fractures who presented to the ED. Demographic characteristics, vital signs, laboratory parameters, and clinical outcomes were recorded. The predictive performance of HALP, mHALP, and mREMS scores for mortality was evaluated using ROC analysis, and the area under the curve (AUC) were compared using the DeLong test.

Results: The mean age of the patients was 78.59±6.81 years, and 63.5% were female. The 30-day mortality rate was 6.3% (n=4). In patients who died, heart rate and mREMS score were significantly higher, while oxygen saturation was significantly lower. ROC analysis showed that the mREMS score had apparent discriminative performance (AUC: 0.981; 95% confidence interval: 0.942-1.00; p<0.001); however, this estimate should be interpreted cautiously given the low number of mortality events (n=4). For mREMS, at a cut-off value of 7.5, sensitivity was 100% and specificity was 84.7%. HALP (AUC: 0.581; p=0.690) and mHALP (AUC: 0.513; p=0.942) scores showed no significant predictive value for mortality. The DeLong test results showed a statistically significant difference between mREMS and mHALP scores (p=0.009).

Conclusion: The mREMS score showed apparent superiority over HALP and mHALP in predicting 30-day mortality; however, these findings should be considered exploratory and require confirmation in larger studies.

Keywords: Hip fractures, aged, mortality, emergency service, hospital, risk assessment

Introduction

Trauma is a leading cause of death worldwide. For individuals aged 5-29, three of the top five causes of death are trauma-related, and 8% of all deaths are trauma-related (1). In the United States, unintentional injuries, a major cause of trauma, rank as the third leading cause of death after heart disease and malignancies (2). In the United States, traumatic injuries are a major cause of both mortality and morbidity. In 2014, there were 30,838,741 non-fatal injury admissions to emergency departments (EDs) in the US.

Hip fractures represent a significant proportion of unintentional injuries. Motor vehicle accidents and falls are common causes of hip fractures. According to Web-based Injury Statistics Query and Reporting System data developed by the Centers for Disease Control and Prevention, falls are among the leading causes of ED visits and hospitalizations for injuries among older adults (3). With increasing life expectancy, the incidence of hip fractures is expected to rise. Consequently, driven in part by osteoporosis, hip fractures remain a major global public health concern, with more than 10 million cases annually worldwide (4). One study predicts that the number of hip fractures will double by 2050 compared to 2018 (5).

Corresponding Author: Abdülaziz Doğan MD, University of Health Sciences Türkiye, Konya City Hospital, Clinic of Emergency Medicine, Konya, Türkiye
E-mail: dr.aazizdogan@outlook.com **ORCID ID:** orcid.org/0000-0002-1306-3938

Cite this article as: Doğan A, Ünal A. Mortality prediction in geriatric hip fracture patients in the emergency department: a comparison of HALP, mHALP, and mREMS scores. Eurasian J Emerg Med. 2026;25: 367-74.

Received: 23.02.2026
Accepted: 21.06.2026
Published: 29.06.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

The Rapid Emergency Medicine score (REMS) is an effective triage score for predicting in-hospital mortality among non-trauma hospital admissions (6). REMS consists of variables such as age, mean arterial pressure (MAP), heart rate, respiratory rate, oxygen saturation, and the glasgow coma scale (GCS). This score was subsequently modified for trauma patients. It is also a simple and accurate predictor of in-hospital mortality among trauma patients. The modified REMS (mREMS) is an adapted version of the REMS score designed to serve as a practical, real-time triage score, in contrast to more complex scores that often require invasive measurements (7).

The hemoglobin, albumin, lymphocyte, and platelet score (HALP), a composite index derived from routine laboratory parameters, was first studied in patients with malignancy and found to be a significant predictor of mortality (8). The HALP score is a useful prognostic biomarker in various malignancies, including myelodysplastic syndromes (9). In another study, it was also found to be a significant predictor of mortality in breast cancer (10). Similarly, the results showed that cancer patients with high HALP scores had a lower long-term risk of death compared to patients with low HALP scores, and this was consistent across all subgroups (11). Similarly, the modified HALP score (mHALP) was a significant predictor of 3-month mortality in patients with acute heart failure. An increased mHALP score was associated with a good prognosis (12).

Two studies in patients with hip fractures examined 90-day and 1-year mortality and showed an inverse relationship between HALP score and risk of death (13,14). In addition, the HALP score was compared with C-reactive protein and the C-reactive protein-albumin ratio, and HALP was found to be a more significant predictor of mortality (14).

In patients with hip fractures, mortality may be influenced by both acute physiological deterioration (measured by mREMS) and the patient's underlying chronic reserve (measured by HALP). This study investigated which marker better predicted short-term mortality. In our literature review, we found no studies comparing HALP, mHALP, and mREMS for predicting mortality in patients presenting with hip fractures. Therefore, we hope to contribute to the literature by prospectively examining the HALP score, the mHALP score, and the mREMS in patients presenting to the ED with hip fractures.

Materials and Methods

Study Design and Setting

This was a single-center prospective observational cohort study conducted in the ED of University of Health Sciences Türkiye, Konya City Hospital, Konya, Türkiye, between June 15, 2025, and December 15, 2025, with follow-up for 30-day mortality. The study population included geriatric patients (aged ≥65 years) who presented to the ED with trauma complaints and were diagnosed with a hip fracture (including femur head, neck, intertrochanteric, or subtrochanteric types) confirmed by clinical and imaging evaluations. The primary outcome measure was 30-day all-cause mortality, rigorously tracked and confirmed using the National Death Registry of the Ministry of Health and, when necessary, supplemented by direct telephone follow-up with patients or their legally authorized representatives. Secondary outcomes, hospital stay length and the comparative predictive performance of the HALP, mHALP, and mREMS scores in estimating these outcomes.

The study protocol was approved by the University of Health Sciences Türkiye, Konya City Hospital Non-Interventional Clinical Research Ethics Committee (decision number: 2025/76, date: 12.05.2025). The research was conducted in strict accordance with the ethical principles outlined in the Declaration of Helsinki. Prior to participation, informed written consent was obtained from all patients included in the study or from their legally authorized representatives.

Participants (Inclusion and Exclusion Criteria)

The study population consisted of geriatric patients (aged ≥65 years) who presented to the ED with traumatic injuries.

Inclusion Criteria: Patients aged ≥65 years and diagnosed with a hip fracture (femur head, neck, intertrochanteric, or subtrochanteric types) confirmed by clinical evaluation and imaging (X-ray or computed tomography).

Exclusion Criteria: Patients brought to the ED in cardiopulmonary arrest, those with a known diagnosis of malignancy, patients with pathological fractures due to tumors or infections, and those discharged at their own request. Additionally, patients with serious metabolic or infectious diseases (e.g., end-stage renal or hepatic failure, sepsis) that could biologically confound HALP and mHALP scores were excluded.

Sample Size and Statistical Power

The sample size was determined by the number of consecutive eligible patients presenting during a predefined six-month study period. This period was selected to ensure feasibility and to capture a representative sample of geriatric hip fracture patients presenting to a high-volume tertiary ED. A total of 63 patients were included in the final analysis.

As this study was designed as an exploratory, hypothesis-generating investigation, no formal a priori sample size calculation was performed. However, considering the primary

objective of evaluating discriminative performance using ROC analysis, a minimum number of outcome events is generally recommended to obtain stable area under the curve (AUC) estimates. In this study, the number of mortality events (n=4) was substantially below commonly suggested thresholds, which limits the precision and reliability of the performance estimates.

A post-hoc assessment based on the observed ROC performance of the mREMS score (AUC=0.981) suggested that the observed effect was statistically detectable; however, given the very low number of mortality events (n=4), the assessment does not overcome the limitations in precision, stability, and the risk of optimism bias. Therefore, the findings should be interpreted as exploratory and require confirmation in larger, adequately powered studies.

To further address the limitations related to sample size, confidence intervals (CIs) and bootstrap resampling (1000 iterations) were used to improve the robustness of the estimates. Despite the large observed effect size, the low number of events may still lead to overestimation of model performance, and the results should be interpreted with caution.

Data Collection and Variables

The study was conducted prospectively. Consecutive sampling was used to minimize selection bias. Vital signs of patients presenting to the ED with trauma were obtained according to standard clinical practices and evaluated by the emergency physician. Patients diagnosed with a hip fracture based on imaging were considered eligible for inclusion. No additional laboratory tests or imaging were performed for the purposes of the study; only data obtained during routine clinical evaluation were used. Data were recorded from patients who provided informed consent. Patients who met the exclusion criteria were excluded from the study. In the included patients, age, gender, GCS, systolic blood pressure (SBP), respiratory rate, heart rate, SpO_2, mREMS score, hemoglobin (g/L), albumin (g/L), lymphocyte (10^9/L), platelet (10^9/L), HALP score, mHALP score, fracture type, surgical treatment status, in-hospital outcome, and 30-day mortality data were recorded.

The prognostic performance of three distinct scoring systems was evaluated:

HALP score: This index reflects the immuno-nutritional status and was calculated as:

HALP=[hemoglobin (g/L) × albumin (g/L) × lymphocyte count (10^9/L)] / platelet count (10^9/L)

mHALP score: To assess variations in predictive accuracy based on modified weightings of nutritional and inflammatory components, the mHALP score was calculated as:

mHALP=[hemoglobin (g/L) × albumin (g/L) × lymphocyte count ($\times10^9$/L) × platelet count ($\times10^9$/L)].

All parameters used in the calculation of HALP and mHALP scores were obtained from routine laboratory measurements at the time of ED admission, and standard SI units were consistently applied to ensure reproducibility. For score calculations, laboratory values originally measured in different units (e.g., g/dL) were converted to consistent SI-equivalent units prior to computation.

mREMS score: The mREMS was calculated based on physiological parameters (age, heart rate, respiratory rate, SBP, SpO_2, and GCS) to provide a standardized snapshot of acute physiological distress upon ED presentation.

Following data collection, patients were grouped according to their 30-day mortality status. The predictive performance of HALP, mHALP, and mREMS scores for mortality was assessed by comparing patients who died with those who survived.

Rationale for Comparison

The rationale for comparing these specific scores lies in the multi-faceted nature of geriatric trauma, where outcomes are determined by both the severity of the acute injury and the patient's underlying physiological reserve. While mREMS is a validated tool for assessing acute physiological instability in the ED, it primarily captures the immediate stress response and does not account for the patient's baseline nutritional status or chronic inflammatory state. Conversely, the HALP and mHALP scores serve as objective indicators of the host's biological resilience and immune-nutritional capacity, both of which are critical determinants of recovery in older adults. By comparing these scores, the study aims to determine whether objective laboratory-based indices can supplement or even outperform traditional physiological scores in identifying high-risk geriatric patients who may otherwise appear stable during initial triage.

Statistical Analysis

Statistical analyses were performed using Jamovi (version 2.7.13). As the study was conducted prospectively with real-time data collection, no missing data were observed; therefore, no imputation was required. The normality of continuous variables was assessed using the Shapiro-Wilk test and was supported by graphical methods. Continuous variables were expressed as mean ± standard deviation or median (interquartile range), as appropriate, while categorical variables were presented as frequencies and percentages.

Group comparisons were performed using Student's t-test or the Mann-Whitney U test, depending on the distribution of the data and homogeneity of variances. Categorical variables were

analyzed using the chi-square test or Fisher's exact test, when appropriate.

Because of the small sample size and the limited number of mortality events (n=4), effect sizes were calculated to complement p-values. Rank-biserial correlation was used for non-parametric comparisons, and effect size measures were not systematically reported for all parametric and categorical variables.

The predictive performance of HALP, mHALP, and mREMS scores for 30-day mortality was evaluated using ROC curve analysis. Internal validation was performed using bootstrap resampling (1000 iterations) to estimate 95% CIs for the AUC. ROC curves were compared using the DeLong test. Optimal cut-off values were determined using the Youden index, and sensitivity, specificity, and predictive values were calculated with 95% CIs. Given the low number of events, the results of these analyses should be interpreted with caution. A p-value<0.05 was considered statistically significant.

In addition, an exploratory logistic regression analysis was performed to assess the association between mREMS score and 30-day mortality. However, due to the very small number of outcome events (n=4), the model was prone to overfitting and to complete separation. Therefore, adjusted effect estimates could not be reliably obtained, and the regression analysis was interpreted descriptively.

Results

A total of 84 patients were initially assessed for eligibility. Twenty-one patients were excluded for the following reasons: malignancy (n=8), treatment refusal (n=3), and severe metabolic or infectious conditions (n=10). Consequently, 63 patients were included in the final analysis.

The mean age of the 63 patients was 78.59±6.81 years, and 63.5% (n=40) were female. Most patients (93.7%, n=59) had at least one comorbid condition at admission. Regarding fracture types, 52.4% (n=33) were intertrochanteric fractures and 47.6% (n=30) were femoral neck fractures. All patients underwent surgical treatment. Most ED admissions resulted in ward hospitalization (96.8%, n=61), while 3.2% (n=2) required intensive care. The median length of hospital stay was 6 (4-10) days. The 30-day mortality rate was 6.3% (n=4). Given the small number of mortality events, subsequent analyses should be interpreted with caution because of limited statistical robustness. Patient demographics, clinical characteristics, and outcome data are presented in Table 1.

In the mortality group, heart rate (102.00 vs. 80.34 beats/min; p<0.001) and mREMS score (9 vs. 6 points; p<0.001) were significantly higher, while oxygen saturation was significantly lower (p=0.036). The length of hospital stay was also significantly longer in this group (p=0.047). No significant differences were observed between groups with respect to age, SBP, respiratory rate, or laboratory parameters, including albumin, hemoglobin, platelet count, lymphocyte count, and HALP and mHALP scores. A comparison of the groups based on 30-day mortality is shown in Table 2.

The distribution of mREMS scores across mortality groups is shown in Figure 1.

ROC analysis, conducted to evaluate the mREMS score's ability to predict 30-day mortality, yielded an AUC of 0.981 (95% CI: 0.942-1.00), which was statistically significant (p<0.001).However, this very high AUC value should be interpreted with caution, as it

Table 1. Baseline demographic, clinical, laboratory characteristics, and hospital outcomes of the study population

Variables	Total (n=63)
Demographic characteristics	
Age (years), mean ± SD	78.59±6.81
Gender, n (%)	
Woman	40 (63.5%)
Man	23 (36.5%)
Clinical features	
Comorbid diseases, n (%)	
Yes	59 (93.7%)
No	4 (6.3%)
Fracture type, n (%)	
Intertrochanteric	33 (52.4%)
Femoral neck	30 (47.6%)
Vital signs	
Systolic blood pressure (mmHg), median (IQR)	130 (119-140)
Heart rate (beats/min), mean ± SD	81.71±12.27
Respiratory rate (/min), median (IQR)	14 (13-15)
Oxygen saturation (%), mean ± SD	94.54±3.01
Laboratory parameters	
Hemoglobin (g/L), mean ± SD	121.7±18.4
Albumin (g/L), mean ± SD	36.13±4.46
Platelet (109/L), mean ± SD	228.83±80.79
Lymphocyte (109/L), median (IQR)	1.17 (0.79-1.67)
Emergency department outcomes	
Admission to service, n (%)	61 (96.8%)
Admission to intensive care unit, n (%)	2 (3.2%)
In-hospital endings	
Hospital stay length (days), median (IQR)	6 (4-10)

SD: Standard deviation, IQR: Interquartile range

Table 2. Comparison of patients based on 30-day mortality status			
Variables	**Survivors (n=59)**	**Deaths (n=4)**	**p-value**
Age (years)*	78.6±6.7	77±4.08	0.634
Vital signs			
Heart rate (beats/min)*	80.34±9.15	102.00±30.16	<0.001
Oxygen saturation (%)*	94.75±2.83	91.50±4.36	0.036
Systolic blood pressure (mmHg)^	130 (120-140)	125 (110-135)	0.470
Respiratory rate (/min)^	14 (13-15)	15 (14-18)	0.178
Laboratory values			
Albumin (g/L)*	36.2±4.4	37.5±4.2	0.529
Hemoglobin (g/L)*	122±18	119±21	0.810
Platelet (109/L)*	229.1±80.5	215.5±95.2	0.513
Lymphocyte (109/L)^	1.18 (0.80-1.68)	1.05 (0.65-1.45)	0.390
Scores			
HALP score^	24.4 (14.6-40.5)	15.5 (13.1-765)	0.602
mHALP score (103/L)^	1206 (698-1733)	869 (763-6656)	0.944
mREMS score^	6 (6-6)	9 (8.75-9)	<0.001
Hospital stay (days)^	6 (4-9.5)	13 (10.25-17.75)	0.047
Values marked with *are presented as mean ± standard deviation, whereas values marked with ^are presented as median (interquartile range)			

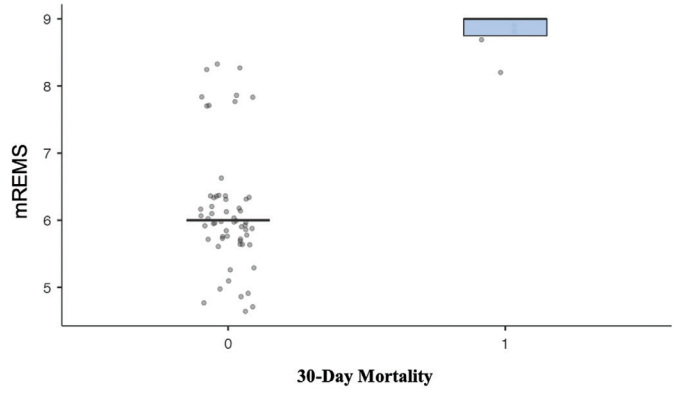


Figure 1. Distribution of mREMS scores according to 30-day mortality status (Box Plot)

mREMS: Modified Rapid Emergency Medicine score

is likely influenced by the extremely small number of outcome events and may reflect overestimation of model performance (optimism bias).

Using the Youden index, the optimal cut-off value for mREMS was determined to be 7.5. Based on this cut-off value, the following classification was made:

Four of 13 patients (30.8%) with an mREMS score of 7.5 or higher died.

No deaths occurred among the 50 patients (0%) with an mREMS score below 7.5.

Based on these findings, the sensitivity of the test at a cut-off value of 7.5 was 100%, and its specificity was 84.7%. The negative predictive value (NPV) was 100% in this study sample. Nevertheless, the observed perfect sensitivity and NPV should be interpreted cautiously, as these estimates are highly unstable and prone to inflation in small samples with very few events.

Univariate logistic regression analysis using the mREMS score failed to produce stable estimates because of complete separation: no mortality events were observed among patients with lower mREMS values. This resulted in inflated coefficients and standard errors, precluding meaningful interpretation of odds ratios. This further highlights the limitations imposed by the low event rate on model-based inference.

In contrast, when the predictive power of HALP and mHALP scores for mortality was examined in the same patient group, no statistically significant discriminatory ability was observed. The AUC for the HALP score was 0.581 (95% CI: 0.185-0.976, p=0.690) and the AUC for the mHALP score was 0.513 (95% CI: 0.171-0.854, p=0.942). Given their limited and non-significant discriminative performance, further diagnostic classification analyses were not pursued for the HALP and mHALP scores (Figure 2).

Pairwise comparisons of ROC curves were performed using the DeLong test. The mREMS score demonstrated significantly greater discriminative performance than the mHALP score (AUC difference=0.4682; 95% CI: 0.1188-0.818; p=0.009).

The AUC difference between mREMS and HALP scores was 0.4004, but it did not reach statistical significance (95% CI:-0.0149-0.816; p=0.059).

No significant difference was observed between HALP and mHALP scores for AUC (AUC difference=0.0678; 95% CI:-0.1052-0.241; p=0.442).

Discussion

This study compared the performance of HALP, mHALP, and mREMS scores in predicting 30-day mortality in geriatric patients presenting to the ED with hip fractures. The findings indicate that the mREMS score shows higher apparent discriminative performance in predicting mortality, whereas HALP and mHALP scores do not significantly contribute to discriminating short-term mortality. These findings suggest that scoring systems incorporating acute physiological parameters may be more suitable for predicting short-term outcomes in this patient population.

The strong performance of the mREMS score in our study can be explained by its components, which reflect acute physiological deterioration. The mREMS score was initially developed to predict mortality in non-surgical patients presenting to the ED, and combining vital signs and age has been shown to improve its prognostic power (6). Later, the mREMS score, adapted and improved for the trauma population, has been reported to perform better than most existing trauma scores in predicting mortality (7). Since a significant portion of hip fracture patients falls within a clinical spectrum where acute trauma physiology and frailty syndrome intersect, a score measuring the deterioration of acute physiological reserve is expected to better predict short-term mortality. The strong association of acute physiological parameters, such as increased heart rate, decreased oxygen saturation, and high mREMS scores, with death in patients in our study also supports this approach.

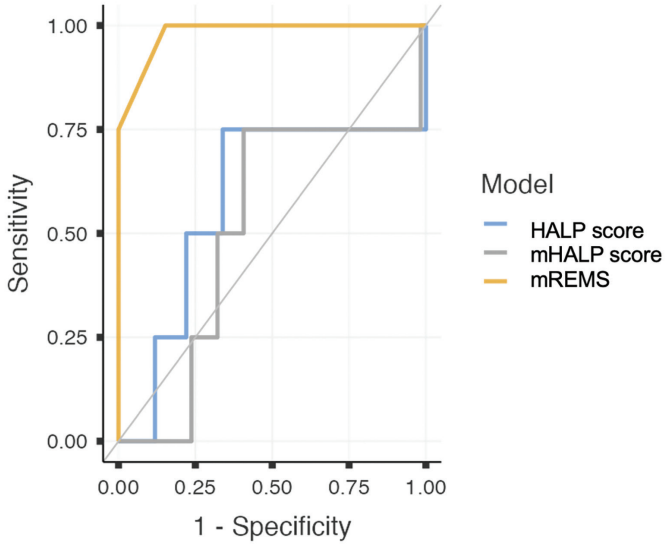


Figure 2. ROC curve analysis for predicting 30-day mortality. The predictive performance of the scoring systems is illustrated. mREMS demonstrated strong discriminative ability, with an AUC of 0.981 (95% CI: 0.942-1.000, p<0.001). HALP and mHALP scores showed lower, non-significant predictive performance, with AUCs of 0.581 (95% CI: 0.185-0.976, p=0.690) and 0.513 (95% CI: 0.171-0.854, p=0.942), respectively

mREMS: Modified Rapid Emergency Medicine score, AUC: Area under the curve, CI: Confidence interval, mHALP: modified hemoglobin, albumin, lymphocyte, and platelet score

In contrast, the finding that HALP and mHALP scores were not significant predictors of short-term mortality is valuable. The HALP score was developed as a composite index that evaluates the patient's inflammatory, immunological, and nutritional status by combining HALP levels (8). In the literature, the HALP score is most often associated with long-term prognosis in patients with malignancy and can serve as an independent prognostic marker (9-11). The common feature of these studies is that the HALP score reflects chronic inflammation, nutritional status, and immune reserve rather than acute physiological instability. Therefore, the limited performance of the HALP score in ED populations, where acute physiological changes are prominent in determining short-term mortality, is biologically plausible.

Similar results have been reported in ED-based studies. A study in patients with acute heart failure showed that the mHALP score had moderate prognostic value. In contrast, the classical HALP score is insufficient to predict early and short-term mortality (12). This finding supports the idea that biochemical composite scores, which do not directly measure physiological deterioration in acute clinical conditions, may be limited in this regard. In our study, the lack of significant differences in HALP and mHALP scores across mortality groups suggests that mortality risk is determined primarily by the acute clinical condition.

Studies evaluating HALP scores in hip fracture populations have focused primarily on medium- and long-term mortality. In large-sample cohort studies, HALP scores have been reported to be independently associated with 90-day and long-term mortality and to improve predictive performance when added to risk models (13). Similarly, in elderly proximal femur fracture patients, HALP scores have been reported to be effective in predicting 30-day and 1-year mortality and to show stronger discrimination compared to other inflammatory markers (14). Several factors may account for the differences between our study and these studies. First, the relatively small sample size and a low number of mortality events may have limited the statistical power, particularly for detecting weaker associations, such as those related to HALP-based indices. Second, since HALP

scores are markers of chronic reserve, their effect may be limited to a short time interval, such as 30 days.

In addition to these findings, recent studies have reported mixed results regarding the predictive value of HALP in hip fracture populations. In a cohort of elderly patients undergoing hemiarthroplasty for femoral neck fractures, the HALP score was identified as an independent predictor of six-month mortality, with moderate discriminative performance (AUC≈0.80) (15). These findings support the role of HALP as a marker of overall physiological reserve and mid-term outcomes.

However, not all studies have demonstrated a consistent association between HALP and short-term mortality. In a more recent ED-based cohort, although HALP scores were numerically lower in non-survivors, no statistically significant association was found with either 30-day or 1-year mortality (16). Instead, factors such as age, hypoalbuminemia, and indicators of acute clinical severity, including requirement for intensive care and prolonged hospitalization, were more strongly associated with mortality.

Taken together, these findings suggest that while HALP may have prognostic value in certain contexts, particularly for longer-term outcomes, its role in predicting early mortality in the emergency setting remains uncertain. This variability across studies further supports the notion that acute physiological parameters may be more relevant for short-term risk assessment in this patient population.

The very high AUC observed for the mREMS score in this study should be interpreted with caution. In small-sample studies, model performance may be overestimated due to optimism bias or overfitting. Although internal validation using bootstrap resampling was performed, external validation in larger, independent cohorts is required to confirm the robustness and generalizability of these findings. However, the high sensitivity and NPV of the mREMS score, at a cut-off of 7.5, suggest that it may be useful for identifying low-risk patients in early clinical settings.

In addition, adjustment for potential confounding factors could not be performed reliably. Although an exploratory logistic regression analysis was attempted, the extremely low number of mortality events (n=4) resulted in complete separation and unstable parameter estimates. Therefore, the independent effects of variables such as comorbidities, frailty, and baseline clinical status could not be evaluated; residual confounding cannot be excluded. Importantly, the absence of a multivariable analytical approach limits the ability to determine whether the observed associations, particularly for the mREMS score, are independent of these clinically relevant factors. As such, the apparent predictive performance of the mREMS score may be partially influenced by unmeasured or uncontrolled confounders.

Another key finding of this study is that scores assessing acute physiological status may be more predictive of short-term mortality than chronic biochemical indices. In geriatric patients with hip fractures, mortality is influenced not only by underlying comorbidities but also by the physiological response to acute trauma. Therefore, the use of easily applicable scores, such as mREMS, in triage and early risk assessment processes can support clinical decision-making. From a clinical perspective, incorporating such tools into ED workflows may facilitate more accurate risk stratification and optimize resource allocation, particularly in high-risk geriatric populations.

Study Limitations

This study has several strengths, including the head-to-head evaluation of acute physiological (mREMS) and immuno-nutritional (HALP, mHALP) indices in a specific geriatric hip fracture cohort. To our knowledge, direct comparisons of these three scores in predicting short-term mortality are limited, making this study a potentially valuable contribution to the literature.

However, several limitations must be acknowledged. First, the single-center design and relatively small sample size (n=63) limit the generalizability of our findings. The low number of mortality events (n=4) significantly constrains the statistical power of the analyses. This may have influenced the stability of the performance estimates, including the high AUC value observed for the mREMS score. In small-sample settings, such findings may be affected by optimism bias or model overfitting.

Additionally, the limited number of events may have reduced the ability to detect statistically significant associations for HALP and mHALP scores, despite their potential clinical relevance. Furthermore, the study primarily relied on univariate analyses and ROC-based discrimination methods, without adjusting for potential confounding variables. Important clinical factors such as comorbidities, frailty status, and baseline functional or physiological conditions could not be incorporated into a multivariable model due to the limited number of outcome events. As a result, the independent predictive value of the evaluated scores could not be determined, and residual confounding may have influenced the observed associations. Therefore, the findings of this study should be interpreted with caution and considered exploratory.

Finally, the lack of external validation is another important limitation. Multicenter prospective studies with larger sample sizes and independent validation cohorts are needed to confirm

the robustness and clinical applicability of these scoring systems in geriatric trauma populations.

Conclusion

This study suggests that the mREMS score appears to demonstrate discriminative performance in predicting short-term mortality and may outperform the HALP and mHALP scores in this patient population. However, these findings should be interpreted with caution due to the extremely limited number of mortality events, which may have led to overestimation of the AUC and sensitivity.

In contrast, HALP and mHALP scores did not show significant predictive value for 30-day mortality. The high sensitivity and NPV of the mREMS score indicate potential clinical utility, particularly for identifying low-risk patients; however, these results should be considered exploratory.

Further validation in larger, adequately powered studies is required.

Ethics

Ethics Committee Approval: The study protocol was approved by the University of Health Sciences Türkiye, Konya City Hospital Non-Interventional Clinical Research Ethics Committee (decision number: 2025/76, date: 12.05.2025). The research was conducted in strict accordance with the ethical principles outlined in the Declaration of Helsinki.

Informed Consent: Written informed written consent was obtained from all patients included in the study or from their legally authorized representatives.

Footnotes

Authorship Contributions

Surgical and Medical Practices: A.D., Concept: A.D., Design: A.D., Data Collection or Processing: A.D., Analysis or Interpretation: A.D., A.Ü., Literature Search: A.D., A.Ü., Writing: A.D., A.Ü.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. World Health Organization. Injuries and violence [Internet]. Geneva: World Health Organization. Available from: https://www.who.int/news-room/fact-sheets/detail/injuries-and-violence
2. Xu J, Murphy SL, Kochanek KD, Arias E. Mortality in the United States, 2024. NCHS Data Brief. 2026:10.15620/cdc/174641.
3. Centers for disease control and prevention (CDC). Web-based injury statistics query and reporting system (WISQARS) [Internet]. Available from: https://www.cdc.gov/injury/wisqars/index.html
4. GBD 2019 fracture collaborators. Global, regional, and national burden of bone fractures in 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019. Lancet Healthy Longev. 2021;2:e580-e92.
5. Sing CW, Lin TC, Bartholomew S, Bell JS, Bennett C, Beyene K, et al. Global epidemiology of hip fractures: secular trends in incidence rate, post-fracture treatment, and all-cause mortality. J Bone Miner Res. 2023;38:1064-75. Epub 2023 May 29.
6. Olsson T, Terent A, Lind L. Rapid emergency medicine score: a new prognostic tool for in-hospital mortality in nonsurgical emergency department patients. J Intern Med. 2004;255:579-87.
7. Miller RT, Nazir N, McDonald T, Cannon CM. The modified rapid emergency medicine score: a novel trauma triage tool to predict in-hospital mortality. Injury. 2017;48:1870-7. Epub 2017 Apr 25.
8. Chen XL, Xue L, Wang W, Chen HN, Zhang WH, Liu K, et al. Prognostic significance of the combination of preoperative hemoglobin, albumin, lymphocyte and platelet in patients with gastric carcinoma: a retrospective cohort study. Oncotarget. 2015;6:41370-82.
9. Gursoy V, Sadri S, Kucukelyas HD, Hunutlu FC, Pinar IE, Yegen ZS, et al. HALP score as a novel prognostic factor for patients with myelodysplastic syndromes. Sci Rep. 2024;14:13843.
10. Jiang T, Sun H, Xue S, Xu T, Xia W, Wang Y, et al. Prognostic significance of hemoglobin, albumin, lymphocyte, and platelet (HALP) score in breast cancer: a propensity score-matching study. Cancer Cell Int. 2024;24:230.
11. Dong J, Jiang W, Zhang W, Guo T, Yang Y, Jiang X, et al. Exploring the J-shaped relationship between HALP score and mortality in cancer patients: a NHANES 1999-2018 cohort study. Front Oncol. 2024;14:1388610.
12. Kocaoglu S, Alatli T. The Efficiency of the HALP Score and the Modified HALP score in predicting mortality in patients with acute heart failure presenting to the emergency department. J Coll Physicians Surg Pak. 2022;32:706-11.
13. Wang Z, Liu H, Liu M. The hemoglobin, albumin, lymphocyte, and platelet score as a useful predictor for mortality in older patients with hip fracture. Front Med (Lausanne). 2025;12:1450818.
14. Vural A, Dolanbay T, Yagar H. Hemoglobin, albumin, lymphocyte and platelet (HALP) score for predicting early and late mortality in elderly patients with proximal femur fractures. PLoS One. 2025;20:e0313842.
15. Tahak F, Yaka H, Kırılmaz A, Kekeç AF, Çolak TS, Özer M. Relationship between mortality and HALP score in femoral neck fractures treated with hemiarthroplasty. Jt Dis Relat Surg. 2025;36:589-95. Epub 2025 Jun 20.
16. Durgun HM, Özkul E. Prognostic value of CCI and HALP score in geriatric femur fractures: one-year emergency department experience. Dicle Med J. 2026;53:77-85.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Letter to the Editor
DOI: 10.4274/eajem.galenos.2026.05882
Eurasian J Emerg Med. 2026;25: 375-6

Inadvertent Intravenous Administration of an Oral Alginate-antacid Suspension: A Wrong-route Medication Error and Strategies for Prevention

Havva Serin Yiğit

Konya Meram State Hospital, Clinic of Emergency Medicine, Konya, Türkiye

Keywords: Medication errors, wrong-route administration, alginate-antacid, patient safety

Dear Editor,

Medication administration errors are among the leading causes of hospital-acquired adverse events, and the intravenous (IV) route carries the highest risk among such errors (1). Although the inadvertent IV administration of preparations intended for the oral or enteral route is classified as a "never event," it continues to be reported in practice; the Pennsylvania Patient Safety Authority has documented cases of serious harm, including one death, attributable to such errors (2). Wrong-route events have been reported following administration of other oral or enteral liquids, including simethicone, which precipitated respiratory distress and cyanosis (3), enteral feeds, and breast milk, which have been associated with microembolism, coagulation activation, and even death (4). To the best of our knowledge, however, no case involving an oral alginate–antacid suspension has previously been reported. We present what we believe to be the first reported case, observed in our emergency department, of inadvertent IV push administration of an oral alginate–antacid suspension containing sodium alginate, sodium bicarbonate, and calcium carbonate.

A 34-year-old man with no known chronic illness presented to the emergency department with heartburn and dizziness accompanying symptoms of an upper respiratory tract infection that had begun one day earlier. He was in good general condition with stable vital signs: blood pressure 115/85 mmHg, heart rate 75 beats/min, oxygen saturation 98% on room air, and temperature 37.5 °C. The ordered treatment consisted of IV paracetamol, an IV esomeprazole push, and an oral alginate-antacid suspension (Gasvin, Deva Holding, İstanbul, Türkiye; sodium alginate 500 mg, sodium bicarbonate 267 mg, and calcium carbonate 160 mg per 10 mL). The product's Summary of Product Characteristics reports neither the pH nor the osmolarity of this oral preparation (5). The oral suspension was drawn up into a parenteral (Luer) syringe and handed to a trainee nurse for administration without the intended oral route being communicated. Assuming that a filled syringe implied IV use, the trainee administered 10 mL of the suspension by IV push. The error was recognized promptly by the nursing staff and the patient's relative, who noted that the medication was normally taken by mouth. The patient was immediately moved to the resuscitation area, the attending physician was notified, and serial electrocardiography (ECG) monitoring was initiated. Apart from a transient, self-limiting ecchymosis at the injection site, no systemic findings were observed in the early period, and the ECG showed a normal sinus rhythm. The National Poison Information Center was consulted and recommended follow-up by toxicology and intensive care, with possible hemodialysis.

Immediately after the event, the patient was hemodynamically stable and asymptomatic apart from an injection-site ecchymosis; serial monitoring at 90 minutes showed stable vital signs (blood pressure 125/80 mmHg, pulse 108 beats/min, temperature 36.5 °C). Arterial blood gas and routine laboratory parameters—including complete blood count, biochemistry, cardiac enzymes, lactate dehydrogenase, D-dimer, and the coagulation profile

Corresponding Author: Havva Serin Yiğit, Konya Meram State Hospital, Clinic of Emergency Medicine, Konya,Türkiye
E-mail: hvvsrnygt@gmail.com **ORCID ID:** orcid.org/0009-0009-1196-2057
Cite this article as: Serin Yiğit H. Inadvertent intravenous administration of an oral alginate-antacid suspension: a wrong-route medication error and strategies for prevention. Eurasian J Emerg Med. 2026;25: 375-6.

Received: 27.05.2026
Accepted: 16.07.2026
Published: 29.07.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

(PT/aPTT/INR)—were within normal limits. A chest radiograph demonstrated no acute cardiopulmonary abnormality. Contrast-enhanced computed tomography, initially considered to evaluate iatrogenic pulmonary microembolism, was deferred owing to concerns about the additional hemodynamic and renal burden of contrast media; no alternative cross-sectional imaging was performed. The patient was transferred to the toxicology intensive care unit of a tertiary urban hospital for further follow-up care.

On admission to intensive care, new symptoms that were not present initially were recorded (back pain, aural burning, and bilateral leg pain), while vital signs remained stable. Fluid therapy, nasal cannula oxygen, and serial blood-gas monitoring were provided. Following a nephrology consultation, preparations for possible hemodialysis were made, but hemodialysis was not required. During the three-day follow-up, no hemolysis, hepatic or renal dysfunction, electrolyte abnormalities, or elevations of cardiac biomarkers developed; clinical and laboratory parameters were within normal limits, and the patient was discharged uneventfully.

Oral suspensions containing sodium alginate, sodium bicarbonate, and calcium carbonate act locally: on contact with gastric acid, they form a viscous gel ("raft") and do not require systemic absorption (6). Their physicochemical properties, however, make the IV route inherently hazardous. The suspension contains insoluble particulate calcium carbonate and is highly viscous, so intravascular injection carries a theoretical risk of vascular obstruction and pulmonary microembolism—the mechanism proposed for respiratory deterioration after IV simethicone and IV enteral feeds, in which microembolism of water-insoluble particles and non-specific activation of coagulation have been implicated (3,4). In addition, the sodium bicarbonate load and alkaline pH raise concerns about hemolysis, metabolic alkalosis, and hypernatremia, while the particulate calcium carbonate content may cause hypercalcemia. These considerations underpinned the poison center's recommendation for intensive monitoring with readiness for hemodialysis, and the initial plan to perform imaging for microembolism. The absence of significant systemic complications in our patient probably reflects the small volume administered (10 mL), immediate recognition of the error, and close monitoring but should not be interpreted as evidence that the IV administration of oral preparations is safe.

Route confusion in patients concurrently prescribed oral and IV medications arises when latent weaknesses in system defenses align (7). In our case, three such factors were evident: an oral suspension was drawn into a parenteral syringe that could physically connect to an IV line; the intended route was not communicated during task hand-off to a trainee nurse; and no barrier—such as an independent double-check or a route-specific connector—prevented the administration from occurring. Notably, the medication had been ordered on paper and via the electronic order-entry (e-order) system, yet the wrong-route administration still occurred, indicating that the vulnerability in this case lay in the administration (execution) step rather than in the prescribing step. The error was recognized only after the dose had been given, promptly and jointly by the nursing staff and the patient's relative, allowing immediate escalation and monitoring. Preventive strategies, therefore, need to target the administration step: physical separation of oral and IV preparations, dispensing oral liquids only in dedicated oral/enteral syringes, independent double-checks before administration, high-alert medication lists, and structured supervision of trainees. Specially designed enteral connectors that cannot be physically connected to IV lines (the ISO 80369-3/ENFit system) offer an additional engineering-based safeguard; however, these systems are not yet in widespread use in our country (4,8).

We report what we believe is the first case of inadvertent IV administration of an oral alginate–antacid suspension—a rare but potentially serious wrong-route medication error. The benign course observed here should not be mistaken for evidence of safety. Beyond individual vigilance, system-level engineering and process safeguards are essential to prevent similar events in emergency department practice.

Footnotes

Financial Disclosure: The author declared that this study received no financial support.

References

1. Sutherland A, Canobbio M, Clarke J, Randall M, Skelland T, Weston E. Incidence and prevalence of intravenous medication errors in the UK: a systematic review. Eur J Hosp Pharm. 2020;27:3-8.
2. Shah-Mohammadi AR, Gaunt MJ. Oral medications inadvertently given via the intravenous route. Pa Patient Saf Advis. 2013;10:85-91.
3. Lozovskytė A, Badaras R, Laubner Sakalauskienė G. Accidental intravenous administration of simethicone in a 4-year-old patient: a case report. Acta Med Litu. 2024;31:149-51.
4. Döring M, Brenner B, Handgretinger R, Hofbeck M, Kerst G. Inadvertent intravenous administration of maternal breast milk in a six-week-old infant: a case report and review of the literature. BMC Res Notes. 2014;7:17.
5. Gasvin 500 mg + 267 mg + 160 mg/10 mL oral suspension. Summary of product characteristics. Deva Holding A.Ş., İstanbul, Türkiye. Marketing authorization no. 2016/256 (first authorized 12 April 2016). Turkish Medicines and Medical Devices Agency (TİTCK). Available from: https://www.titck.gov.tr/kubkt
6. Leiman DA, Riff BP, Morgan S, Metz DC, Falk GW, French B, et al. Alginate therapy is effective treatment for GERD symptoms: a systematic review and meta-analysis. Dis Esophagus. 2017;30:1-9.
7. Kuitunen S, Niittynen I, Airaksinen M, Holmström AR. Systemic causes of in-hospital intravenous medication errors: a systematic review. J Patient Saf. 2021;17:e1660-8.
8. Ethington S, Volpe A, Guenter P, Simmons D. The lingering safety menace: a 10-year review of enteral misconnection adverse events. Nutr Clin Pract. 2024;39:1251-8.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Case Report
DOI: 10.4274/eajem.galenos.2026.65807
Eurasian J Emerg Med. 2026;25: 377-81

Multi-vessel Coronary Cameral Fistulas: The Role of Multimodality Imaging in Guiding Conservative Management

Mehmet Kış[1], Çisem Oktay[1], Mustafa Mahmut Barış[2]

[1]Dokuz Eylül University Faculty of Medicine, Department of Cardiology, İzmir, Türkiye
[2]Dokuz Eylül University Faculty of Radiology, İzmir, Türkiye

Abstract

Coronary cameral fistulas (CCF) are rare anomalies that can lead to significant morbidity due to abnormal coronary shunting, which may result in coronary steal, volume overload, and aneurysmal dilatation of the affected segments. Current guidelines recommend that the management strategy—whether conservative follow-up, percutaneous intervention, or surgical closure—be individualized by a multidisciplinary team with expertise in congenital heart disease, while carefully considering the fistula's origin, drainage anatomy, haemodynamic impact, patient age, and comorbidities. We report a rare case of a 66-year-old woman with multiple CCFs involving all three major coronary arteries, which were incidentally detected during elective coronary angiography. To evaluate potential myocardial ischemia secondary to coronary steal, cardiac magnetic resonance imaging was performed, revealing no pathological signal abnormalities. Following a multidisciplinary heart team evaluation, a conservative management strategy was adopted due to the absence of significant functional impairment. The patient has remained asymptomatic throughout a two-year follow-up period. This case underscores the critical importance of functional assessment over anatomical complexity in guiding clinical management, highlighting that multimodality imaging is essential for identifying patients who can safely avoid unnecessary invasive interventions.

Keywords: Coronary cameral fistula, right ventricular drainage, multimodality imaging, conservative management

Introduction

Krause first described the concept of coronary artery fistula (CAF) in 1865 (1). In contemporary series, the prevalence of coronary fistulas is reported to range from 0.08% to 0.3% among patients undergoing coronary angiography. This rate increases to approximately 0.9% with the use of computed tomography (CT) or magnetic resonance angiography (2). In contrast, the incidence of coronary cameral fistulas (CCF) in the general population is exceedingly low, estimated at approximately 0.0002% (3). CAFs may be either congenital or acquired, with approximately 90% of cases considered to be congenital (4).

CCF is defined as a rare congenital anomaly characterised by an abnormal connection between a coronary artery and a cardiac chamber. Under normal conditions, coronary arteries originate from the aorta and terminate in the myocardial bed; however, in the presence of CCF, blood flow deviates from its physiological course and shunts into a cardiac chamber. The right ventricle is reported to be the cardiac cavity most commonly drained by CCFs (5). In contrast, coronary artery-left ventricle fistulas are extremely rare, accounting for only approximately 2-3% of all CCF cases (6). This extremely low incidence of CCF means that evidence is largely limited to case reports; and consequently, each case requires a personalised, meticulous assessment and management strategy to address potential haemodynamic burden and ischaemic complications.

Current guidelines recommend that the choice between conservative management, percutaneous intervention, or surgical closure should be individualized by a multidisciplinary team comprising cardiologists and cardiac surgeons experienced

Corresponding Author: Mehmet Kış MD, Dokuz Eylül University Faculty of Medicine, Department of Cardiology, İzmir, Türkiye
E-mail: drmehmet.kis@gmail.com **ORCID ID:** orcid.org/0000-0003-0775-8992

Cite this article as: Kış M, Oktay Ç, Barış MM. Multi-vessel coronary cameral fistulas: the role of multimodality imaging in guiding conservative management. Eurasian J Emerg Med. 2026;25: 377-81.

Received: 01.06.2026
Accepted: 18.07.2026
Published: 29.07.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

in congenital heart disease, taking into account the anatomical characteristics of the fistula, its haemodynamic significance, and the patient's clinical profile (4,7). We report a rare case of multiple CCF involving all three major coronary arteries that were incidentally detected during elective coronary angiography.

Case Report

A 66-year-old woman with no prior history of cardiac disease but with known diagnoses of left adrenal adenoma and primary hyperaldosteronism (Conn's syndrome), presented to our cardiology outpatient clinic with typical angina. Her medical history was notable for hypertension, hyperlipidemia, and active smoking (50-pack-year history). She was receiving spironolactone and dexamethasone for the adrenal adenoma and telmisartan for hypertension.

On presentation, the electrocardiogram demonstrated a normal sinus rhythm at a rate of 72 beats per minute. Transthoracic echocardiography revealed a normal left ventricular ejection fraction and mild mitral, tricuspid, and aortic regurgitation. The maximum aortic valve velocity (V_{max}) was 1.3 m/s, and the estimated pulmonary artery pressure was 25 mmHg. The estimated Qp/Qs ratio was 1.1, with preserved right ventricular size and function, and no right-sided chamber enlargement or volume overload. Given her family history of coronary artery disease, comorbidities, and typical angina, she underwent coronary computed tomography (CCT) angiography. CCT angiography demonstrated an abnormal vascular course with focal contrast opacification in the distal extension of the acute marginal branch of the right coronary artery (RCA), adjacent to the interventricular septum (IVS) at the mid- ventricular level, with drainage into the right ventricle (Figure 1). In addition, a thin septal branch originating from the distal left anterior descending (LAD) artery was observed coursing within the myocardium and was associated with focal contrast opacification and communication with the right ventricular cavity, consistent with a coronary-cameral fistula (Figure 1). Furthermore, one of the distal branches of the circumflex (CX) artery was found to course within the inferior wall at the apical level and drain into the right ventricle via the IVS. Collectively, these findings indicated the presence of multiple CCF involving all three major coronary arteries that drained into the right ventricle.

In addition, focal mid-myocardial contrast enhancement consistent with adipose tissue was identified at the left ventricular apex. A cardiac magnetic resonance (CMR) viability study was recommended to evaluate the possibility of ischemia related to coronary steal and was subsequently performed. Invasive coronary angiography demonstrated that the LAD and CX arteries were tortuous and ectatic, whereas the RCA was tortuous, non-dominant, and exhibited 20-30% proximal segment stenosis. CCF were identified in all three coronary arteries (Figures 2,3). Acetylsalicylic acid, metoprolol, and atorvastatin were added to the patient's existing medical therapy. CMR imaging was performed mainly to assess viability and scarring. Stress perfusion CMR was not performed. Late gadolinium enhancement imaging demonstrated no evidence of ischemic myocardial scarring in the sagittal, axial, and coronal planes, corresponding to Figures 4A, 4B, and 4C, respectively.

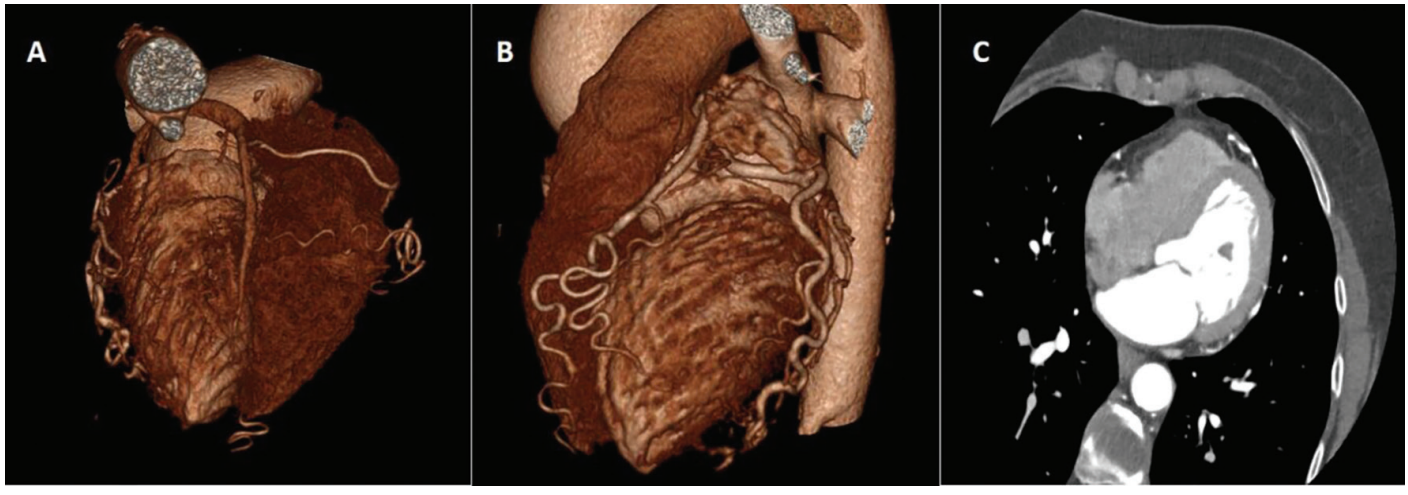


Figure 1. Coronary cameral fistulas on 3D computed tomography (CT) coronary angiography

Three-dimensional computed tomography (CT) coronary angiography demonstrating coronary cameral fistulas

(A) Fistula originating from the right coronary artery (B) fistula originating from the left anterior descending artery (LAD) (C) axial contrast-enhanced CT image showing an abnormal vascular structure consistent with an LAD-originating fistula

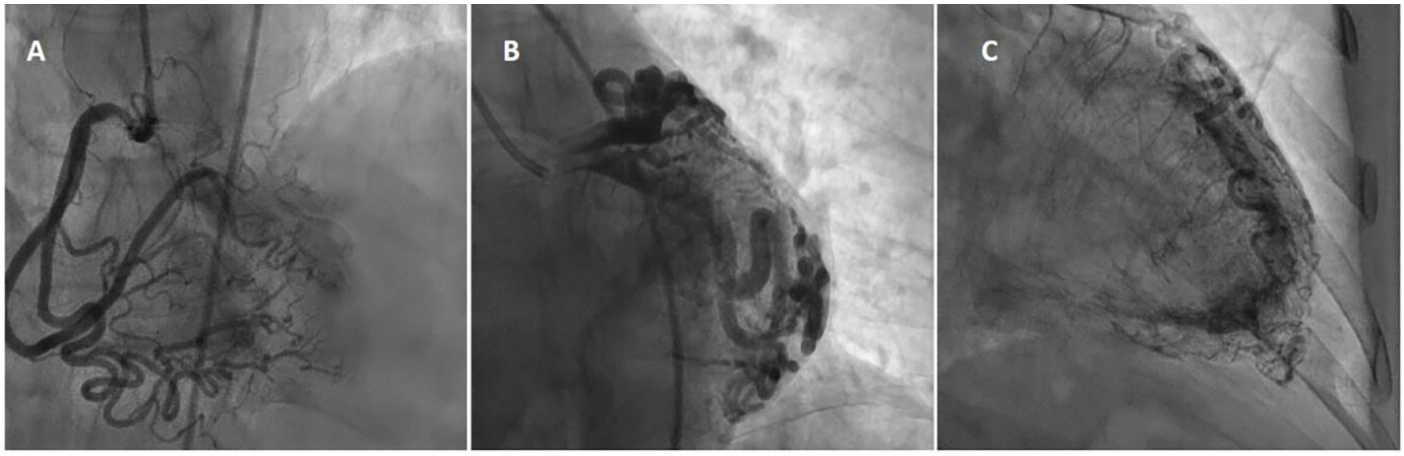


Figure 2. Early ventricular filling in conventional coronary angiography

In the coronary angiography images, early filling consistent with coronary artery fistulas is observed in the right, circumflex, and left coronary arteries, respectively (A, B, and C)

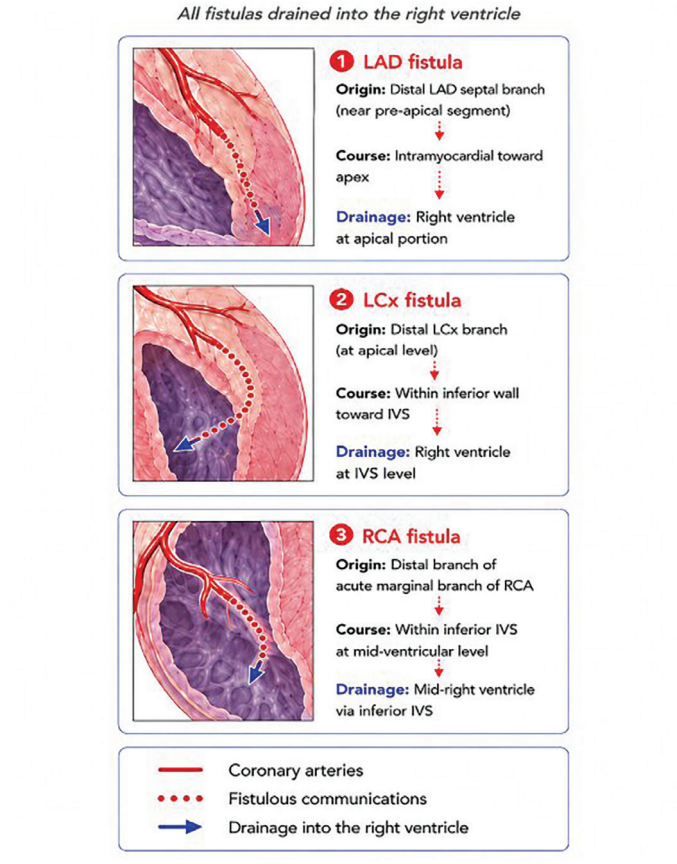


Figure 3. Schematic illustration of triple-vessel coronary-cameral fistulas

The illustration summarizes fistulous communications arising from the left anterior descending artery, circumflex artery, and right coronary artery, all draining into the right ventricle

LAD: Left anterior descending artery, LCx: Left circumflex artery, RCA: Right coronary artery

T1 mapping was obtained, demonstrating myocardial T1 values within normal limits. T2-weighted short-axis images showed no evidence of myocardial edema. Overall, left ventricular myocardial viability was preserved, with no evidence of ischemic scar, edema, or regional wall motion abnormality. Exercise treadmill testing was negative for inducible ischemia; the patient completed stage III of the Bruce protocol and achieved 10 METs without exercise-induced angina or ischemic ST-segment changes.

Following the initiation of medical therapy with acetylsalicylic acid, atorvastatin, and metoprolol, the patient's anginal symptoms completely resolved. The multidisciplinary heart team favored conservative follow-up with medical therapy in this individual patient, given the absence of heart failure, inducible ischemia on exercise testing, haemodynamically significant left-to-right shunting, right-sided volume overload, and pulmonary hypertension. The patient has been followed regularly for two years and remains asymptomatic on medical therapy.

Discussion

A CAF is a vascular malformation characterized by an abnormal communication between a coronary artery and a cardiac chamber or major vessel and may be either congenital or acquired. The fundamental pathophysiological mechanism involves the diversion of high-pressure coronary arterial blood away from the physiological low-pressure arteriolar and capillary network supplying the myocardium toward a low-resistance venous compartment (8).

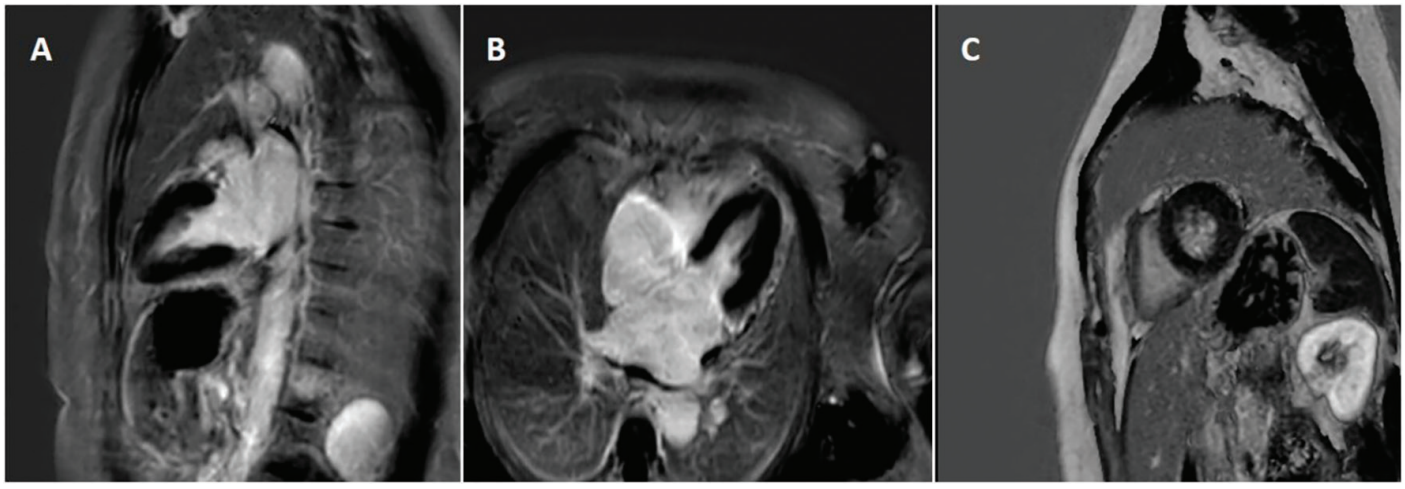


Figure 4. Cardiac magnetic resonance viability imaging

Late gadolinium enhancement images showed no evidence of ischemic myocardial scar. Sagittal (A), axial/transverse (B), and coronal (C) planes are shown

Although many coronary fistulas remain asymptomatic, symptomatic cases may present with heart failure, endocarditis, conduction abnormalities, or myocardial ischemia due to the coronary steal phenomenon, most commonly manifesting as angina and dyspnea (9). The haemodynamic impact of the fistula is primarily determined by its origin, course, and drainage site. Therefore, accurate assessment of the anatomical origin and termination of the fistula is essential to identify perfusion abnormalities distal to the fistula related to coronary steal (8).

Coronary steal is recognized as a key mechanism underlying exertional angina, particularly in conditions associated with increased myocardial oxygen demand, such as physical exertion. Small-diameter fistulas generally follow a benign natural course and often do not require treatment. In contrast, medium and large fistulas are associated with clinically significant long-term complications, including angina, myocardial infarction, arrhythmias, heart failure, and endocarditis. Accordingly, the presence of symptoms, fistula-related complications, or a haemodynamically significant shunt constitutes the main indications for percutaneous closure or surgical repair in patients with CAF (4,10). Current guidelines recommend individualized assessment by a multidisciplinary team experienced in congenital and acquired heart disease to determine the relative merits of medical therapy, percutaneous closure, or surgical repair in patients with CAF. However, because no clear consensus exists regarding the optimal timing or modality of closure, management should be guided by a case-specific risk–benefit assessment that balances anatomical complexity and procedural risk against the anticipated natural history and haemodynamic significance of the fistula (4,7).

In a surgical series of 46 patients, the predominant preoperative symptoms were angina and heart failure. In the same study, the incidence of postoperative myocardial infarction, attributed to reduced flow in the dilated coronary artery proximal to the site of fistula closure, reached 11%. Furthermore, long-term survival was significantly lower compared with an age-matched reference population (8). In another reported case, a 31-year-old patient with a giant aneurysmal fistula originating from the sinus node artery of the RCA and draining into the superior vena cava underwent successful surgical management, including aneurysm resection, aortocoronary bypass using a saphenous vein graft, and fistula closure. The postoperative course was uneventful, and the patient was discharged in good clinical condition (11).

In our case, the preserved left ventricular systolic function, the absence of significant volume overload or pulmonary hypertension on echocardiography, and the lack of obstructive coronary artery disease on invasive coronary angiography suggest a limited haemodynamic impact of the fistulas. CMR imaging showed no myocardial scar or regional wall motion abnormality. Exercise treadmill testing demonstrated no objective evidence of inducible ischemia: no anginal symptoms or ischemic ST-segment changes were observed.

Although a complex network of fistulas involving all three major coronary arteries was present, the potential morbidity and mortality associated with surgical or transcatheter closure were considered high relative to the patient's current symptom burden and clinical stability. Following a multidisciplinary heart-team evaluation, conservative follow-up with medical therapy was favored for this patient in the absence of symptomatic heart

failure, inducible ischemia on exercise testing, haemodynamically significant left-to-right shunting, right-sided volume overload, or pulmonary hypertension. This case highlights that extensive anatomical involvement does not necessarily translate into haemodynamic significance, emphasizing the importance of functional assessment in guiding management decisions.

Unlike previous reports of triple-vessel CCF, which have mainly described left ventricular drainage or symptomatic presentations, the present case involved fistulous drainage from all three major coronary arteries into the right ventricle. Importantly, this anatomical complexity was not associated with significant shunt burden, right-sided volume overload, pulmonary hypertension, or adverse events during two years of conservative follow-up.

Conclusion

Congenital CCF are rare malformations of the coronary arteries and are infrequently detected on coronary angiography. Management strategies vary widely, ranging from close clinical follow-up in asymptomatic patients with haemodynamically insignificant lesions to transcatheter closure in anatomically suitable cases and to surgical repair in selected patients.

Ethic

Informed Consent: Written informed consent was obtained from the patient.

Data Sharing Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Footnotes

Author Contributions

Surgical and Medical Practices: M.K., Concept: M.K., Ç.O., Design: M.K., M.M.B., Data Collection or Processing: M.K., Ç.O., M.M.B., Analysis or Interpretation: M.K., Ç.O., M.M.B., Literature Search: M.K., Ç.O., M.M.B., Writing: M.K., Ç.O.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Krause W. On the origin of an accessory coronary artery from the pulmonary artery. Z Ratl Med. 1865;24:225-7.
2. Lim JJ, Jung JI, Lee BY, Lee HG. Prevalence and types of coronary artery fistulas detected with coronary CT angiography. AJR Am J Roentgenol. 2014;203:W237-43.
3. Seon HJ, Kim YH, Choi S, Kim KH. Complex coronary artery fistulas in adults: evaluation with multidetector computed tomography. Int J Cardiovasc Imaging. 2010;26(Suppl 2):261-71.
4. Stout KK, Daniels CJ, Aboulhosn JA, Bozkurt B, Broberg CS, Colman JM, et al. 2018 AHA/ACC guideline for the management of adults with congenital heart disease. Circulation. 2019;139:e698-e800.
5. Chia BL, Chan AL, Tan LK, Ng RA, Chiang SP. Coronary artery-left ventricular fistula. Cardiology. 1981;68:167-79.
6. Pérez-Pomares JM, de la Pompa JL, Franco D, Henderson D, Ho SY, Houyel L, et al. Congenital coronary artery anomalies: a bridge from embryology to anatomy and pathophysiology. Cardiovasc Res. 2016;109:204-16.
7. Baumgartner H, De Backer J, Babu-Narayan SV, Budts W, Chessa M, Diller GP, et al. 2020 ESC guidelines for the management of adult congenital heart disease. Eur Heart J. 2021;42:563-645.
8. Said SM, Burkhart HM, Schaff HV, Connolly HM, Phillips SD, Suri RM, et al. Late outcome of repair of congenital coronary artery fistulas—a word of caution. J Thorac Cardiovasc Surg. 2013;145:455-60.
9. Lim ECN, Lim CED. Coronary artery fistula presenting with coronary steal syndrome: the significance of clinical vigilance and second opinions. Clin Case Rep. 2025;13(9):e70874.
10. Narh JT, Zahid E, Shivaraj K, Sahni S, Kariyanna PT, Khan A. Steal and strain: a case of coronary artery fistula presenting with coronary steal syndrome. Respir Med Case Rep. 2020;31:101301.
11. Vural U, Aglar AA. A giant sinus node artery fistulizing to superior vena cava presenting with steal phenomenon: an unusual case report. Cardiovasc Surg Interv. 2018;5:43-6.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.59752
Eurasian J Emerg Med. 2026;25: 382-7

The Importance of Blood Parameters for Prognosis Prediction in Status

Turgut Dolanbay[1], Abdussamed Vural[2], Mustafa Oğuz Cumaoğlu[2], Ömer Yüceer[2], Gizem Çakan[2], Durmuş Ayan[3], Naci Ömer Alayunt[4]

[1]Malatya Turgut Özal University Faculty of Medicine, Department of Emergency Medicine, Malatya, Türkiye
[2]Niğde Ömer Halisdemir University Faculty of Medicine, Niğde Ömer Halisdemir University Training and Research Hospital, Department of Emergency Medicine, Niğde, Türkiye
[3]Niğde Ömer Halisdemir University Faculty of Medicine, Niğde Ömer Halisdemir University Training and Research Hospital, Department of Medical Biochemistry, Niğde, Türkiye
[4]Siirt University Medical Faculty of Medicine, Department of Medical Biochemistry, Siirt, Türkiye

Abstract

Aim: To compare blood parameters between patients with status epilepticus (SE) and those with non-SE epileptic seizures at presentation to the emergency department, and to evaluate the prognostic value of immature granulocytes (IG).

Materials and Methods: This retrospective study included adult patients presenting with seizures. Hematological and biochemical parameters, including C-reactive protein (CRP), white blood cell count (WBC), blood gas parameters, IG, and IG% were analyzed.

Results: A total of 634 patients were screened; 68 hospitalized patients who met the inclusion criteria were analyzed (SE: n=37; non-SE: n=31). Significant differences between groups were observed for WBC, lactate, pH, bicarbonate (HCO_3^-), base deficit, CRP, IG, and IG% ($p<0.05$ for all). An IG cut-off value of 0.045 predicted clinical outcomes with 95.8% sensitivity and 97.7% specificity.

Conclusion: IG and IG% are promising, rapid, and cost-effective inflammatory biomarkers with high diagnostic accuracy in SE. Their use may support early risk stratification, guide clinical decision-making, and improve timely intervention in emergency settings.

Keywords: Status epilepticus, immature granulocyte, prognosis

Introduction

Epilepsy is a chronic neurological disorder characterized by recurrent, unprovoked seizures and affects approximately 1% of the global population (1). Status epilepticus (SE) is a life-threatening neurological emergency defined as a prolonged seizure or recurrent seizures without recovery of consciousness between episodes. Although traditionally defined as seizures lasting more than 30 minutes, current clinical practice considers seizures lasting longer than 5 minutes as SE, given the low likelihood of spontaneous termination and the increased risk of neuronal injury (2-4). Early recognition and timely intervention are critical, as delays in treatment are associated with increased morbidity and mortality.

The pathophysiology of SE involves complex mechanisms, including sustained neuronal hyperexcitability, metabolic derangements, hypoxia, and systemic inflammatory responses. Several laboratory parameters such as lactate, leukocyte count, and acid–base disturbances have been investigated as potential biomarkers reflecting disease severity and clinical outcomes (5-7). However, these markers are often nonspecific and may not adequately support early risk stratification in emergency settings.

Immature granulocytes (IG), which represent early-stage granulocytic precursors released from the bone marrow, have recently emerged as markers of systemic inflammation. IG levels can be rapidly and reliably measured using automated hematology analyzers as part of routine complete blood count

Corresponding Author: Turgut Dolanbay MD, Malatya Turgut Özal University Faculty of Medicine, Department of Emergency Medicine, Malatya, Türkiye
E-mail: turgutdolanbay@hotmail.com **ORCID ID:** orcid.org/0000-0002-4092-1192
Cite this article as: Dolanbay T, Vural A, Cumaoğlu MO, Yüceer Ö, Çakan G, Ayan D, et al. The ımportance of blood parameters for prognosis prediction in status. Eurasian J Emerg Med. 2026;25: 382-7.

Received: 23.05.2026
Accepted: 21.07.2026
Published: 13.08.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

testing. Previous studies have demonstrated the prognostic value of IG in various clinical conditions, including sepsis, acute pancreatitis, and intracerebral hemorrhage (8-10). Despite these findings, the role of IG in neurological emergencies, particularly in SE, remains poorly understood.

We hypothesized that elevated IG and IG% on admission to the emergency department may serve as early predictive biomarkers of disease severity and need for urgent intervention in patients with SE.

To the best of our knowledge, no previous study has specifically evaluated the diagnostic and prognostic utility of IG in differentiating SE from non-SE epileptic seizures. Therefore, this study aims to compare hematological and biochemical parameters between these patient groups and investigate the potential role of IG and IG% as accessible, rapid, and clinically useful biomarkers in the early assessment of SE.

Materials and Methods

This retrospective, single-center observational study was conducted after approval by the Niğde Ömer Halisdemir Ethics Committee. The study included patients admitted to the Emergency Department of Niğde Ömer Halisdemir Training and Research Hospital between 01.12.2021 and 01.12.2023 with a diagnosis of SE (ICD-10 code G41.9), and patients admitted with epileptic seizures without an SE diagnosis (ICD-10 code G40.9). Demographic characteristics of the patients, including age and sex, were recorded. In addition, seizure type and changes in seizure-related blood parameters were analyzed. The study was conducted in accordance with ethical protocols. Our study began after approval by the Niğde Ömer Halisdemir Ethics Committee (protocol number: 2023/129, date: 04.01.2024).

The blood and biochemical parameters included: C-reactive protein (CRP), white blood cell count (WBC), blood gas parameters (lactate, pH, bicarbonate, base excess), and IG count and percentage. IG and IG% values were measured using the Sysmex XN-1000 automated hematology analyzer, which uses fluorescence flow cytometry, to ensure standardized and reproducible measurements.

Inclusion Criteria

All patients aged ≥18 years who presented to the adult emergency department of Niğde Ömer Halisdemir Training and Research Hospital between 01.12.2021 and 01.12.2023 with a diagnosis of SE or epileptic seizure and whose clinical and laboratory data were fully accessible were included in the study.

Exclusion Criteria

Patients younger than 18 years and those with comorbid conditions such as diabetes mellitus, hypertension, pregnancy, renal failure, oncologic diseases, and chronic liver disease were excluded. In addition, patients without a confirmed diagnosis of epileptic seizure or SE and those with incomplete data were excluded from the analysis.

Statistics Analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, Version 27.0 (IBM Corp., Armonk, NY, USA). The normality of data distribution was assessed using the Kolmogorov-Smirnov test. Continuous variables with normal distribution were expressed as mean ± standard deviation, while non-normally distributed variables were presented as median and interquartile range (IQR: 25^{th}-75^{th} percentile). Categorical variables were expressed as frequencies and percentages.

Comparisons between categorical variables were performed using Pearson's chi-square test or Fisher's exact test. Continuous variables were compared using Student's t-test or Mann–Whitney U test, as appropriate. Correlation analyses between continuous variables were conducted, and correlation coefficients with corresponding significance levels were reported. ROC curve analysis was performed to evaluate the diagnostic performance of IG and IG% in distinguishing SE from non-SE epilepsy.

A p-value <0.05 was considered statistically significant.

Results

A total of 634 patients were evaluated. A total of 68 patients who met the criteria for evaluation were hospitalized. 36 of these patients had convulsive SE, and 1 had the non-convulsive SE subtype. 37 patients were diagnosed with SE. In addition, the number of patients admitted with epileptic seizures was 31.

There was no statistically significant difference between the SE group and epilepsy according to gender as a result of the χ^2 test ($p=0.641$). The mean age of the patients was 42.25±20.11 years and the median age (IQR: 25-75) was 37 (22.25-59.75). Of the patients, 48.5% were female and 51.5% were male.

Statistically significant differences between SE and non-SE epilepsy groups were found in WBC count and percentage, lactate, pH, bicarbonate (HCO_3^-), base deficit, CRP, and IG. This relationship is given in Table 1.

In the correlation analysis, statistically significant positive correlations of moderate to strong magnitude were found between IG and WBC ($r=0.537$, $p<0.001$) and between IG and neutrophils ($r=0.475$, $p<0.001$), and between IG% and WBC

(r=0.368, p=0.002) and between IG% and neutrophils (r=0.319, p=0.008) (Table 2).

ROC analysis was performed for the number and percentage of IGs with respect to clinical outcome (intubated vs non-intubated patients), and the area under the curve was reported (Figure 1). Optimal cut-off values for IG number and percentage were determined using the Youden index (sensitivity + specificity-1). A cut-off value of 0.045 for IG count provides 95.8% sensitivity and 97.7% specificity. For the IG percentage, a cut-off value of 0.45 yields 95.8% sensitivity and 96.8% specificity. This relationship is presented in Table 3.

Discussion

Epilepsy is a neurological disorder characterized by recurrent seizures. SE is a medical emergency characterized by prolonged epileptic seizures. The biochemical differences between these two conditions often involve changes during and after seizures. However, the information on this issue remains incomplete, and research is ongoing. In a study on IG, which is the subject of research in the diagnosis and treatment of neurological diseases, it was reported that the number of IG effectively showed 30-day mortality in ischemic stroke patients (10). In another study, it was reported that IG number affected the course of poor prognosis in spontaneous intracerebral hemorrhage (11). In addition, a recent study reported that IG count was an effective predictor of hospital mortality in spontaneous intracerebral hemorrhage (12). In our study, IG associated with the immune response provided important data, especially for differentiating SE from epileptic seizures.

In this study, IG count, IG%, neutrophil count, neutrophil % levels differed significantly between groups. Our correlation analysis showed that these inflammatory markers were positively correlated, suggesting a potential link between inflammation severity and disease progression.

Table 1. Comparison of hematological and biochemical parameters between SE and non-SE groups

Variables	Group	N	Mean	SD	Mean rank	p*
WBC	SE	37	13.04	6.48	41.62	<0.001
	Non-SE epilepsy	31	8.57	2.93	26.02	
Lactate	SE	37	5.64	6.10	40.05	0.011
	Non-SE epilepsy	31	2.10	1.70	27.82	
pH	SE	37	7.30	0.14	29.42	0.02
	Non-SE epilepsy	31	7.39	0.04	40.56	
Bicarbonate	SE	37	19.55	5.02	28.01	0.003
	Non-SE epilepsy	31	22.95	1.96	42.24	
Base excess	SE	37	-3.30	7.55	29.07	0.013
	Non-SE epilepsy	31	0.15	2.82	40.98	
CRP	SE	37	23.02	20.56	44.45	0.027
	Non-SE epilepsy	31	14.67	13.01	28.34	
IG	SE	37	0.07	0.05	42.50	<0.001
	Non-SE epilepsy	31	0.02	0.008	24.95	
IG%	SE	37	0.56	0.24	42.39	<0.001
	Non-SE epilepsy	31	0.25	0.006	25.08	

SE: Status epilepticus, SD: Standard deviation, WBC: White blood count, CRP: C-reactive protein, IG: Immature gyranolocyte, *Mann-Whitney U test

Table 2. Correlation analysis of WBC, neutrophil count, IG, and IG%

Correlations (total n= 68)		Correlation coefficient	p*
IG	WBC	0.537**	<0.001
	Neutrophil	0.475**	<0.001
IG%	WBC	0.368**	0.002
	Neutrophil	0.319**	0.008

WBC: White blood count, IG: Immature gyranolocyte, *Pearson, ** Correlation is significant at the 0.01 level (2-tailed)

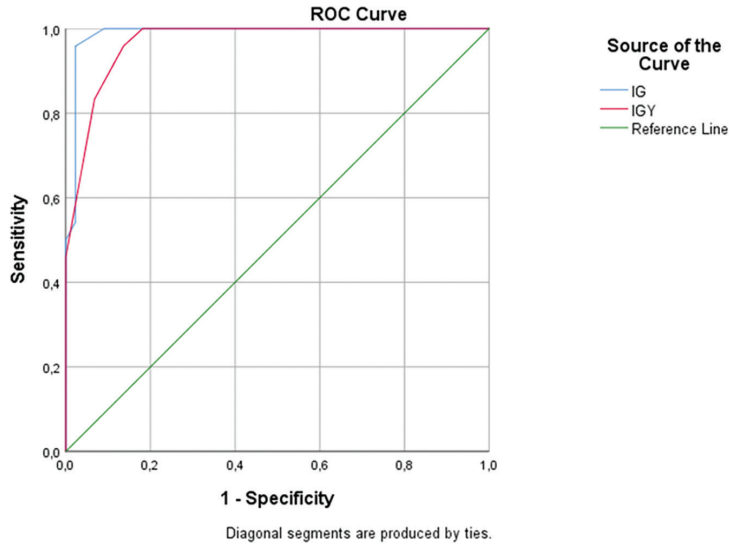


Figure 1. ROC curve graph for IG count and IG%
IG: Immature granulocytes, IGY: Immunoglobulin Y

Table 3. AUC, sensitivity, and specificity of IG count and IG% at optimal cut-off values

Factor	AUC	95% CI (lower-upper)	Sensitivity	Specificity	Youden index	cut-off
IG	IG	0.965-1.000	0.958	0.977	0.936	0.045
IG%	0.968	0.933-1.000	0.958	0.968	0.864	0.4500

AUC: Area under the curve, IG: Immature gyranolocyte, CI: Confidence interval

The elevation of IG during SE may be explained by several underlying pathophysiological mechanisms. Prolonged neuronal hyperactivity can trigger a systemic inflammatory response, leading to cytokine release and activation of the immune system. In addition, seizure-related hypoxia and metabolic stress may stimulate bone marrow activity, resulting in the premature release of granulocytic precursors into the peripheral circulation. Furthermore, increased catecholamine levels during seizures may contribute to leukocyte mobilization. Oxidative stress and neuronal injury may further amplify inflammatory pathways, leading to elevated levels of IG and IG%. These mechanisms provide a plausible biological explanation for the association between IG levels and disease severity observed in our study (13,14).

In the literature, there are findings indicating that inflammatory mediators are released by peripheral immune and brain cells in epileptic seizure processes (15). The significant difference in CRP levels observed in this study suggests that CRP may be used in the differential diagnosis between SE and epilepsy. In one study, the potential use of leukocytosis to differentiate non-epileptic from epileptic seizures was evaluated, and leukocytosis was found to begin approximately 2 hours after the seizure. It was reported that there was a significant difference between epileptic and non-epileptic seizures (16).

A positive correlation was found between IG count and intubated patients in acute respiratory distress syndrome due to coronavirus disease 2019 (COVID-19) (17). In our study, IG in and IG% were determined to be biomarkers with high sensitivity and specificity in intubated patients. Oxidative stress increases during seizures, resulting in neuronal damage and inflammation; consequently, the number and rates of IG, which is an indicator of inflammation, increase. IG, which physicians can measure easily and early, may serve as an early biomarker to guide the decision to intubate. A hemogram is a rapid, low-cost test. Therefore, IG count, other hemogram parameters, and immune response-related markers that can be easily calculated from these parameters will contribute to the differential diagnosis. However, to generalize the findings of this study to large populations, prospective, well-designed studies with large sample sizes that account for time of seizure onset should be conducted. The biggest limitation of this study is that blood tests analyzed immediately after admission to the emergency department were performed only once.

It has been reported in the literature that it is important to analyze lactate, anion gap and ammonia levels at the onset of epileptic seizures (18,19). However, these tests are more costly than a hemogram.

SE may increase lactate levels by affecting energy metabolism and may cause changes in base deficit, bicarbonate, and pH values. Changes in neurotransmitters, such as serotonin and dopamine, can be observed during seizures. These changes are associated with the frequency and severity of seizures. Furthermore, imbalances in neurotransmitters such as gamma-aminobutyric acid and glutamate are associated with epileptic seizures. These imbalances may be more pronounced during SE. During seizures, increased oxidative stress intercellular can be observed. This can lead to cell damage and inflammation. Inflammation and immune-system responses triggered by seizures may lead to changes at the biochemical level, and prolonged seizure duration may exacerbate hypoxia and explain differences in blood gas parameters. These changes may differ between individuals and between seizure types. It is important to remember that epilepsy and SE are complex conditions, and it is natural that the biochemical mechanisms are not yet fully understood. Further research is needed in this area. Treatment and intervention should generally be individualized depending on the type of seizure, the patient's condition and general health (19-21).

The ideal prognostic markers of SE are those that are easily accessible to clinicians at the bedside and provide high sensitivity and specificity in predicting SE-related outcomes. These will provide information on the likelihood of SE outcomes, including progression to refractoriness, functional outcomes, and mortality. Such prognostic markers are essential, as they enable individualization of treatment and aid decisions regarding treatment intensity. No ideal prognostic marker exists for SE, and we are currently limited to the SE markers mentioned above. Despite their good negative predictive value, the positive predictive value of these scores is poor, which prevents their use in decisions to discontinue life-sustaining treatment (20,21). The data obtained in our study will allow the inclusion of clinical, genetic, epigenetic, metabolic, inflammatory, and structural biomarkers in scoring systems that will provide more accurate prognoses.

Our findings show that IG count and % IG levels, in particular, are significant in differentiating SE and epileptic seizures, since they are markers related to the immune response. IG has also shown promising results in COVID-19 and other diseases (21,22). Therefore, we suggest that inflammation and elevated inflammatory markers may help physicians in the early prediction of patients' prognosis and overall condition. In our study, intubated patients with poor general condition had higher IG and IG%.

Study Limitations

This study has several limitations that should be acknowledged. First, its retrospective and single-center design may limit the generalizability of the findings and introduce the potential for selection and information bias. Second, only hospitalized patients were included in the analysis, which may have resulted in a selection bias toward more severe clinical presentations.

Third, IG and IG% levels were measured only once on admission to the emergency department, and the lack of serial measurements precluded assessment of temporal changes in these parameters. In addition, IG levels may vary with the timing of blood sampling relative to seizure onset, which could not be standardized in this study.

Finally, multivariate analysis could not be performed because the limited number of outcome events (e.g., intubation) precluded assessment of IG and IG% as independent predictors. Therefore, the findings should be interpreted with caution, and prospective, multicenter studies with larger sample sizes and repeated measurements are needed to validate these findings.

Conclusion

We observed that inflammatory markers are elevated during epileptic seizures compared with normal reference ranges. Furthermore, we found IG and IG% in epileptic seizures are inflammatory markers with high sensitivity and specificity for predicting early intubation and may guide clinical management and early intervention.

Ethics

Ethics Committee Approval: The study was conducted in accordance with ethical protocols. Our study began after approval by the Niğde Ömer Halisdemir Ethics Committee (protocol number: 2023/129, date: 04.01.2024).

Informed Consent: This retrospective, single-center observational study was conducted after approval by the Niğde Ömer Halisdemir Ethics Committee.

Footnotes

Authorship Contributions

Surgical and Medical Practices: T.D., A.V., M.O.C., D.A., N.Ö.A., Concept: T.D., D.A., Design: T.D., A.V., Data Collection or Processing: D., A.V., M.O.C., D.A., N.Ö.A., Analysis or Interpretation: T.D., M.O.C., Ö.Y., G.Ç., Literature Search: T.D., Ö.Y., G.Ç., Writing: T.D., M.O.C., Ö.Y., G.Ç., D.A., N.Ö.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Uppal P, Cardamone M, Lawson JA. Outcomes of deviation from treatment guidelines in status epilepticus: A systematic review. Seizure. 2018;58:147-53.
2. Purwien L, Schubert-Bast S, Kieslich M, Ronellenfitsch MW, Merker M, Czabanka M, et al. Trends and differences in status epilepticus treatment of children and adults over 10 years: a comparative study of medical records (2012-2021) from a university hospital in Germany. CNS Drugs. 2023;37:993-1008.
3. Coppler PJ, Elmer J. Status epilepticus: a neurologic emergency. Crit Care Clin. 2023;39:87-102.
4. Lee SK. Diagnosis and treatment of status epilepticus. J Epilepsy Res. 2020;10:45-54.
5. Chapman MG, Smith M, Hirsch NP. Status epilepticus. Anaesthesia. 2001;56:648-59.
6. Cherian A, Thomas SV. Status epilepticus. Ann Indian Acad Neurol. 2009;12:140-53.
7. Bambal G, Çakıl D, Ekici F. Mechanisms of the occurrence of epilepsy. Konuralp Medical Journal, 2011;3:42-5.
8. Bedel C, Korkut M, Avcı A, Uzun A. Immature granulocyte count and percentage as new predictors of mortality in patients with upper gastrointestinal bleeding. Indian J Crit Care Med. 2020;24:794-8.
9. Korkut M, Bedel C, Selvi F. Are immature granulocytes and derivatives early predictors of acute appendicitis and acute complicated appendicitis in adults?. Formos J Surg. 2020;53:123-7.
10. Bedel C, Korkut M, Selvi F, Zortuk Ö. The immature granulocyte count is a new predictor of the 30-day mortality in intracerebral haemorrhage patients: preliminary study. Indian J Neurosurg. 2021;10:114-20.
11. Kocatürk İ, Gülten S. Immature granulocyte and other markers in prediction of mortality in spontaneous intracerebral hemorrhage. J Health Sci Med. 2023;6:368-73.
12. Claassen J, Albers D, Schmidt JM, De Marchis GM, Pugin D, Falo CM, et al. Nonconvulsive seizures in subarachnoid hemorrhage link inflammation and outcome. Ann Nurol. 2014;75:771-81.
13. Pender RA, Losey TE. A rapid course through the five electrographic stages of status epilepticus. Epilepsia. 2012;53:193-5.
14. Ansari-Lari MA, Kickler TS, Borowitz MJ. Immature granulocyte measurement using the Sysmex XE-2100. Relationship to infection and sepsis. Am J Clin Pathol. 2003;120:795-9.
15. Li Y, Matzka L, Flahive J, Weber D. Potential use of leukocytosis and anion gap elevation in differentiating psychogenic nonepileptic seizures from epileptic seizures. Epilepsia Open. 2019;4:210-5.
16. Alay GH, Tatlisuluoglu D, Bulut K, Fikri BI, Oztas A, Turan G. The relationship between immature granulocyte count and mortality in ARDS Due to COVID-19. Niger J Clin Pract. 2022;25:1301-7.
17. Doğan EA, Ünal A, Ünal A, Erdoğan Ç. Clinical utility of serum lactate levels for differential diagnosis of generalized tonic-clonic seizures from psychogenic nonepileptic seizures and syncope. Epilepsy Behav. 2017;75:13-7.
18. Shah AK, Shein N, Fuerst D, Yangala R, Shah J, Watson C. Peripheral WBC count and serum prolactin level in various seizure types and nonepileptic events. Epilepsia. 2026;42:1472-5.
19. Vezzani A, Maroso M, Balosso S, Sanchez MA, Bartfai T. IL-1 receptor/Toll-like receptor signaling in infection, inflammation, stress and neurodegeneration couples hyperexcitability and seizures. Brain Behav Immun. 2011;25:1281-9.
20. Semmlack S, Tschudin-Sutter S, Widmer AF, Valença M, Rüegg S, Marsch S, et al. Independent impact of infections on the course and outcome of status epilepticus: a 10-year cohort study. J Neurol. 2016;263:1303-13.
21. Milenkovic J, Djindjic B, Djordjevic B, Stojiljkovic V, Stojanovic D, Petrovic S, et al. Platelet-derived immuno-inflammatory indices show best performance in early prediction of COVID-19 progression. J Clin Lab Anal. 2022;36:e24652.
22. Karakulak S, Narcı H, Ayrık C, Erdoğan S, Üçbilek E. The prognostic value of immature granulocyte in patients with acute pancreatitis. Am J Emerg Med. 2021;44:203-7.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.98105
Eurasian J Emerg Med. 2026;25: 388-93

Serum Semaphorin as a Novel Diagnostic and Prognostic Biomarker in Acute Myocardial Infarction

Sedat Özbay[1], Orhan Özsoy[1], Tansu Gençer[1], Mustafa Ayan[1], Cihan Bedel[2]

[1]Sivas Numune Hospital Clinic of Emergency Medicine, Sivas, Türkiye
[2]University of Health Sciences Türkiye, Antalya Training and Research Hospital, Clinic of Emergency Medicine, Antalya, Türkiye

Abstract

Aim: Acute myocardial infarction (AMI) remains a leading cause of morbidity and mortality worldwide. The identification of novel biomarkers reflecting the inflammatory and thrombotic mechanisms involved in acute coronary syndromes may improve diagnostic accuracy and risk stratification. Semaphorins are a family of proteins involved in immune regulation, vascular biology, and platelet activation, suggesting a potential role in cardiovascular disease. This study aimed to investigate serum semaphorin levels in patients with AMI and to evaluate their potential diagnostic and prognostic value.

Materials and Methods: In this prospective case-control study, 62 participants were enrolled: 31 patients diagnosed with AMI and 31 age- and sex-matched control subjects. Among the AMI patients, 19 had ST-segment elevation myocardial infarction (STEMI) and 12 had non-STEMI (NSTEMI). Serum semaphorin levels were measured using an ELISA. Clinical, demographic, and angiographic data were recorded and analyzed.

Results: The mean serum semaphorin level was 0.21±0.12 in the AMI group and 0.26±0.17 in the control group, with no statistically significant difference between groups (p=0.556). Semaphorin levels were also comparable between male and female participants (0.23±0.14 vs. 0.26±0.18, p=0.662). Subgroup analysis demonstrated mean semaphorin levels of 0.21±0.16 in NSTEMI patients and 0.21±0.07 in STEMI patients. Although semaphorin levels tended to be lower in AMI patients than in the control group, no significant difference was observed among the control, NSTEMI, and STEMI groups(p=0.081).

Conclusion: Serum semaphorin levels were not significantly associated with the presence or subtype of AMI in the present study. These findings suggest limited diagnostic utility of circulating semaphorin levels in AMI. Further multicenter studies with larger sample sizes are needed to clarify the role of semaphorins in the pathophysiology and clinical assessment of acute coronary syndromes.

Keywords: Acute myocardial infarction, semaphorin, biomarker, STEM, NSTEMI, acute coronary syndrome; cardiovascular disease

Introduction

Acute myocardial ınfarction (AMI), commonly known as a heart attack, is a critical cardiac event caused by the sudden blockage of blood flow in one or more coronary arteries, leading to myocardial ischemia and necrosis (1). It remains a leading cause of morbidity and mortality worldwide, with significant prevalence in both developed and developing countries (2,3). AMI includes both ST-segment elevation myocardial infarction (STEMI) and non-ST-segment elevation myocardial infarction (NSTEMI), both of which are managed with similar therapeutic approaches (4).

The relationship between biomarker levels in patients with AMI and their clinical outcomes is multifaceted, involving various biomarkers that reflect different pathophysiological processes. Elevated levels of biomarkers such as NT-proBNP, cTNT, and GDF-15 have been shown to predict heart failure (HF) following AMI, with NT-proBNP demonstrating the highest predictive value for post-AMI HF (5,6). Similarly, BNP and troponin I levels are significantly correlated with adverse outcomes such as left ventricular dysfunction, arrhythmias, and mortality in acute coronary syndrome (ACS) patients, indicating their utility in risk stratification (7,8). Novel biomarkers like sST2, suPAR, and H-FABP

Corresponding Author: Cihan Bedel MD, University of Health Sciences Türkiye, Antalya Training and Research Hospital, Clinic of Emergency Medicine, Antalya, Türkiye
E-mail: cihanbedel@hotmail.com **ORCID ID:** orcid.org/0000-0002-3823-2929

Received: 03.06.2026
Accepted: 21.07.2026
Published: 13.08.2026

Cite this article as: Özbay S, Özsoy O, Gençer T, Ayan M, Bedel C. Serum semaphorin as a novel diagnostic and prognostic biomarker in acute myocardial infarction. Eurasian J Emerg Med. 2026;25: 388-93.

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

are also elevated in AMI patients and correlate with clinical parameters such as ejection fraction and inflammatory markers, suggesting their potential role in assessing cardiac function and predicting outcomes (9). Although semaphorins were initially identified as axon guidance proteins involved in nervous system development, they are now known to have significant effects on the immune system, vascular biology, platelet activation, and inflammation. Specifically, Semaphorin 3A (Sema3A) and Semaphorin 4D (Sema4D) have been found to be associated with endothelial function, atherosclerotic plaque stability, and platelet aggregation. Sema4D is reported to exhibit pro-inflammatory and prothrombotic effects, while Sema3A is known to possess anti-inflammatory and vascular protective properties (10,11).

This study aimed to investigate the diagnostic and prognostic value of serum semaphorin levels in patients with AMI and evaluate the relationship between these levels and clinical, biochemical, and angiographic parameters.

Materials and Methods

This study will be conducted prospectively, observationally, and under controlled conditions on patients presenting to the Emergency Department of the Sivas Cumhurıyet University Health Sciences Research Ethics Committee. The study protocol will be designed in accordance with the principles of the Declaration of Helsinki, and approval will be obtained from the (Relevant Institution) Clinical Research Ethics Committee before data collection begins. Written informed consent will be obtained from all participants (or their immediate family members) included in the study. This study was conducted in accordance with the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee Sivas Cumhurıyet University Health Sciences Research (protocol number: 2025/07-19, date: 10.07.2025).

The study will include consecutive patients who present with chest pain or equivalent symptoms, such as shortness of breath or syncope, and who are diagnosed with AMI according to the Third Universal Definition of Myocardial Infarction. Patient groups: STEMI group: patients with new ST segment elevation in at least two consecutive leads on electrocardiography (ECG) (approximately 2 mm in men, 1.5 mm in women; 1 mm in other leads) or newly developed left bundle branch block and positive cardiac troponin (cTn) kinetics. NSTEMI group: patients with ischemic symptoms and/or dynamic ST-T changes (ST depression, T wave inversion) on ECG with elevated and/or decreased cTn levels, without ST segment elevation. Control group: patients similar in age and gender, with similar cardiovascular risk factors but without coronary artery disease (CAD). Individuals with similar symptoms (e.g., non-cardiac chest pain) in whom normal coronary arteries are detected on angiography or in whom CAD is ruled out by objective tests.

Inclusion Criteria:

- Age ≥ 18 years,
- Admission within the first (12/24) hours from chest pain onset,
- Willingness to participate and provision of signed informed consent.

Exclusion Criteria

To eliminate potential confounding factors that directly influence systemic inflammatory, angiogenic, and immune pathways, patients who meet any of the following criteria will be excluded:

- Active systemic infection, sepsis, or known chronic autoimmune/inflammatory diseases,
- Active malignancy or history of recent chemotherapy/radiotherapy,
- Severe renal impairment (estimated glomerular filtration rate <30 mL/min/1.73 m2 or end-stage hepatic failure,
- History of prior myocardial infarction, coronary artery bypass grafting, or percutaneous coronary intervention,
- Presentation with cardiogenic shock or requiring cardiopulmonary resuscitation at admission.

Patients' ages, genders, myocardial ınfarction thrombolysis scores (TIMI), and angiographic intervention results will be recorded on standard forms. Peripheral venous blood samples will be collected from patients in the AME group upon admission to the emergency department (0 hours, before treatment or angiography). A single basal blood sample will be collected from each healthy volunteers in the control group. Blood samples will be collected in gel-free biochemistry tubes (or EDTA tubes) and centrifuged at 3000 rpm for 10 minutes. The obtained serum/plasma samples will be separated from their cellular components, aliquoted into Eppendorf tubes, and stored in an ultra-low freezer at -80° C until analysis. Semaphorin analysis: human [specific semaphorin type, e.g., semaphorin 3A] levels in the stored samples will be determined using commercially available quantitative ELISA kits specific for human semaphorin according to the manufacturer's instructions. Measurements will be performed using a blinded method (absorbance wavelength: 450 nm) on a microplate reader by a laboratory specialist unaware of the patients' clinical status. Patients' TIMI scores and the extent and severity of CAD after coronary angiography were evaluated by experienced interventional cardiologists.

Statistical Analysis

Data analysis was performed using SPSS version 26.0 (IBM Corp., Armonk, NY, USA). The normality of continuous variables was assessed using the Shapiro-Wilk test. Continuous variables were expressed as mean ± standard deviation for normally distributed data or median (minimum-maximum) for non-normally distributed data. Comparisons between two independent groups were performed using the Student's t-test for normally distributed variables and the Mann-Whitney U test for non-normally distributed variables. Categorical variables were expressed as frequencies and percentages, and compared using the chi-square test. A p value <0.05 was considered statistically significant.

Result

A total of 62 participants were included in the study, comprising 31 patients with AMI and 31 control subjects. Among the AMI patients, 19 were diagnosed with STEMI and 12 with NSTEMI. Coronary angiographic evaluation demonstrated varying degrees of coronary artery stenosis and occlusion among patients with AMI, whereas no significant CAD was detected in the control group.

The mean serum semaphorin level was 0.21±0.12 in the AMI group and 0.26±0.17 in the control group. No statistically significant difference was observed between patients and controls (p=0.556) (Table 1). When serum semaphorin levels were analyzed according to sex, the mean concentration was 0.23±0.14 in males and 0.26±0.18 in females. This difference was not statistically significant (p=0.662). Subgroup analysis according to clinical diagnosis demonstrated mean semaphorin levels of 0.26±0.17 in the control group, 0.21±0.16 in patients with NSTEMI, and 0.21±0.07 in patients with STEMI. Although semaphorin levels tended to be lower in both AMI subgroups than in controls, differences among the three groups did not reach statistical significance (p=0.081).

Table 1. Comparison of serum semaphorin levels according to clinical outcomes and disease severity

Group	**Mean ± SD**	**p**
Patient	0.21±0.12	0.556
Control	0.26±0.17	
Gender		
Male	0.23±0.14	0.662
Female	0.26±0.18	
Population		
Control (n=31)	0.26±0.17	0.081
NSTEMI (n=12)	0.21±0.16	
STEMI (n=19)	0.21±0.07	

NSTEMI: non ST-segment elevation myocardial infarction, STEMI: ST-segment elevation myocardial infarction, SD: Standard deviation

Discussion

This study investigated the diagnostic and prognostic value of serum semaphorin levels in patients with AMI. The principal finding was that serum semaphorin concentrations did not differ significantly between patients with AMI and healthy controls. Furthermore, semaphorin levels were comparable between STEMI and NSTEMI patients and showed no significant association with sex or disease subtype. Although semaphorin concentrations tended to be lower in patients with AMI than in controls, these differences did not reach statistical significance. These findings suggest that circulating semaphorin levels have limited diagnostic utility for distinguishing AMI patients from controls in our study population. The absence of a significant association may be related to the relatively small sample size, the biological variability in semaphorin expression, or the timing of blood sampling during the acute phase of myocardial infarction.

The diagnosis and prognosis of AMI have significantly advanced with the identification and utilization of various biomarkers. cTn, particularly high-sensitivity troponins, are considered the gold standard for diagnosing myocardial injury because of their high sensitivity and specificity, which allow early detection of AMI (12,13). In addition to troponins, creatine kinase MB (CK-MB) and myoglobin are traditional markers used in AMI diagnosis, although their specificity is lower compared to troponins (14,15). Novel biomarkers such as heart-type fatty acid-binding protein and matrix metalloproteinases have shown promise in enhancing diagnostic accuracy (16). For prognostic purposes, N-terminal B-type natriuretic peptide (NT-proBNP) and C-reactive protein are valuable predictors of cardiovascular outcomes, including HF and mortality (17). Recent studies have identified additional biomarkers like growth differentiation factor-15 (GDF-15), tumor necrosis factor-related apoptosis-inducing ligand receptor 2, and fibroblast growth factor 23 as significant predictors of mortality and HF in AMI patients (18). Inflammatory markers such as interleukin-6 and myeloperoxidase also contribute to risk stratification by reflecting underlying inflammatory processes (19). Integrating multiple biomarkers, including novel biomarkers such as microRNAs and galectin-3, into a multi-biomarker approach can enhance diagnostic precision and provide a comprehensive risk assessment (20). However, challenges such as the high cost of testing and potential interference from chronic conditions like renal insufficiency limit the routine clinical application of some novel biomarkers (21). Despite these challenges, the use of artificial intelligence and biosensors in conjunction with biomarkers holds promise for improving early diagnosis and personalized treatment strategies in AMI (22). Overall, while traditional biomarkers remain crucial, the exploration and validation of novel biomarkers continue to evolve.

Although previous studies have demonstrated altered semaphorin expression in several inflammatory and cardiovascular disorders, our study did not identify a significant difference in serum semaphorin levels between patients with AMI and healthy controls. This discrepancy may reflect differences in disease mechanisms, the specific semaphorin isoforms measured, sample size, patient characteristics, or the timing of blood sampling. It is also possible that circulating semaphorin concentrations are influenced by dynamic inflammatory responses that are not adequately captured by a single measurement obtained at hospital admission. Therefore, while semaphorins remain biologically plausible candidates, our findings do not support their routine use as diagnostic biomarkers for AMI. In systemic sclerosis (SSc), Sema3A levels were found to be significantly lower in patients compared to healthy controls, with even lower levels observed in those with major vascular involvements such as digital ulcers and pulmonary arterial hypertension, suggesting its role in vasculopathy and as a potential biomarker for vascular complications in SSc (24,25). Similarly, in systemic lupus erythematosus (SLE), Sema3A levels were significantly reduced in patients, particularly those with active disease, correlating negatively with disease activity indices, indicating its potential as a marker for SLE activit (26,27). In the context of retinal vein occlusion, increased Sema3A levels in aqueous humor were associated with macular edema and ganglion cell degeneration, suggesting a pathological role in these conditions (28). Furthermore, in patients undergoing hemodialysis, Sema3A levels were significantly lower compared to controls and decreased further during dialysis sessions, indicating its potential involvement in kidney function and bone remodeling (29). In reproductive health, elevated Sema3A levels were observed in women with diminished ovarian reserve who responded better to controlled ovarian stimulation, suggesting a role in fertility and reproductive competence (30). On the other hand, Sema4D levels were elevated in HF patients, correlating with inflammatory markers and suggesting its utility as a biomarker for acute HF (31). In periodontal disease, increased Sema4D levels were noted in patients with periodontitis, which decreased following non-surgical periodontal therapy, indicating its role in periodontal inflammation (32). Collectively, these studies underscore the diverse roles of semaphorins in various diseases, highlighting their potential as diagnostic and prognostic biomarkers across different pathological conditions (33).

In AMI, biomarkers such as troponin, CK-MB, and B-type natriuretic peptide are commonly used for diagnosis and risk stratification. However, most of these markers are elevated as a result of myocardial damage (34). Semaphorins, on the other hand, may provide additional information about the pathophysiology of the disease as they may be involved in early stages of processes such as inflammation, platelet activation, and endothelial dysfunction. This suggests that semaphorins may be biomarkers that can be used not only in diagnosis but also in prognostic assessment (35). In our study, we planned to investigate the relationship between semaphorin levels and both TIMI score and coronary angiography findings. The TIMI score is a widely used risk assessment tool for predicting short-term mortality and major cardiac events in patients with ACS. The correlation between semaphorin levels and the TIMI score may reveal the value of these molecules for clinical risk stratification. Similarly, semaphorins can be considered biochemical indicators of atherosclerotic burden if they correlate with the extent and complexity of CAD.

Study Limitations

This study has limitations. First, since the study was conducted at a single center, the generalizability of the results to different populations and healthcare settings may be limited. Second, the relatively small sample size may reduce statistical power and make it difficult to establish significant relationships, especially in subgroup analyses. In this study, semaphorin levels were measured only at the time of admission; serial measurements were not performed. Therefore, the relationship between temporal changes in semaphorin levels and the development of myocardial infarction, receipt of reperfusion therapy, and clinical outcomes could not be evaluated. Furthermore, because only certain semaphorin subtypes were examined, the possible effects of other semaphorin family members could not be investigated.

While the extent and severity of CAD were determined by angiographic evaluation, advanced imaging techniques such as intravascular ultrasonography or optical coherence tomography (OCT) were not used. Therefore, plaque morphology and vulnerability-related characteristics could not be evaluated in detail. Although individuals with active infection, malignancy, and chronic inflammatory disease were excluded from the study, it was not possible to fully control for all biological and environmental factors that could affect semaphorin levels. Furthermore, since long-term follow-up data were not available, the effects of semaphorin levels on long-term mortality, recurrent myocardial infarction, and the development of HF could not be evaluated. Despite these limitations, our study is one of the few studies investigating the relationship between semaphorin levels and clinical, biochemical, and angiographic parameters in AMI, and may therefore contribute to the literature. Future studies with larger sample sizes, multiple centers, and long-term follow-up are needed to confirm these findings.

Conclusion

In this prospective case-control study, serum semaphorin levels did not differ significantly between patients with AMI and healthy controls, nor were they associated with AMI subtype. Although semaphorin concentrations showed a trend toward lower values in patients with AMI, the differences were not statistically significant. These findings suggest that circulating semaphorin levels have limited diagnostic value as standalone biomarkers for AMI in the studied population. Given the relatively small sample size and single-center design, larger multicenter studies with serial measurements and evaluation of specific semaphorin isoforms are warranted to further clarify the potential role of semaphorins in the pathophysiology and clinical assessment of AMI.

Ethics

Ethics Committee Approval: This study was conducted in accordance with the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee Sivas Cumhurıyet University Health Sciences Research (protocol number: 2025/07-19, date: 10.07.2025).

Informed Consent: This is prospective case-control study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: S.Ö., O.Ö., Concept: S.Ö., T.G., Design: S.Ö., O.Ö., T.G., M.A., C.B., Data Collection or Processing: S.Ö., T.G., M.A., Analysis or Interpretation: S.Ö., O.Ö., C.B., Literature Search: S.Ö., O.Ö., T.G., M.A., C.B., Writing: S.Ö., O.Ö., T.G., M.A., C.B.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA, et al. Fourth universal definition of myocardial infarction (2018). Eur Heart J. 2019;40:237-69.
2. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, et al. Heart disease and stroke statistics-2021 update: a report from the American Heart Association. Circulation. 2021;143:e254-e743.
3. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global burden of cardiovascular diseases and risk factors, 1990-2019. J Am Coll Cardiol. 2020;76:2982-3021.
4. Collet JP, Thiele H, Barbato E, Barthélémy O, Bauersachs J, Bhatt DL, et al. 2020 ESC guidelines for acute coronary syndromes. Eur Heart J. 2021;42:1289-367.
5. Omland T, de Lemos JA, Sabatine MS, Christophi CA, Rice MM, Jablonski KA, et al. Prognostic value of N-terminal pro-B-type natriuretic peptide after myocardial infarction. Circulation. 2007;115:1656-64.
6. Wollert KC, Kempf T, Wallentin L. Growth differentiation factor 15 as a biomarker in cardiovascular disease. Clin Chem. 2017;63:140-51.
7. Morrow DA, Cannon CP, Jesse RL, Newby LK, Ravkilde J, Storrow AB, et al. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: clinical characteristics and utilization of biochemical markers in acute coronary syndromes. Circulation. 2007;115:e356-75.
8. O'Donoghue ML, Morrow DA, Cannon CP, Jarolim P, Desai NR, Sherwood MW, et al. Multimarker risk stratification in patients with acute myocardial infarction. J Am Heart Assoc. 2016;5:e002586.
9. Fan Y, Li Y, Chen Y, Zhao YJ, Liu LW, Li J, et al. Comprehensive metabolomic characterization of coronary artery diseases. J Am Coll Cardiol. 2016;68:1281-93.
10. Kumanogoh A, Kikutani H. Immunological functions of the neuropilins and plexins as receptors for semaphorins. Nat Rev Immunol. 2013;13:802-14.
11. Worzfeld T, Offermanns S. Semaphorins and plexins as therapeutic targets. Nat Rev Drug Discov. 2014;13:603-21.
12. ndoval Y, Jaffe AS. Using high-sensitivity cardiac troponin T for acute cardiac care. Am J Med. 2017;130:1358-1365.e1.
13. Chapman AR, Lee KK, McAllister DA, Cullen L, Greenslade JH, Parsonage W, et al. Association of high-sensitivity cardiac troponin ı concentration with cardiac outcomes in patients with suspected acute coronary syndrome. JAMA. 2017;318:1913-24.
14. Apple FS, Collinson PO; IFCC Task force on clinical applications of cardiac biomarkers. Analytical characteristics of high-sensitivity cardiac troponin assays. Clin Chem. 2012;58:54-61.
15. Thygesen K, Mair J, Mueller C, Huber K, Weber M, Plebani M, et al. Recommendations for the use of natriuretic peptides in acute cardiac care: a position statement from the study group on biomarkers in cardiology of the ESC working group on acute cardiac care. Eur Heart J. 2012;33:2001-6.
16. Glatz JF, Reneman RS. Fatty acid-binding proteins in cardiovascular disease. Physiol Rev. 2006;86:1205-40.
17. Braunwald E. Biomarkers in heart failure. N Engl J Med. 2008;358:2148-59.
18. Wallentin L, Hijazi Z, Andersson U, Alexander JH, De Caterina R, Hanna M, et al. Growth differentiation factor 15, a marker of oxidative stress and inflammation, for risk assessment in patients with atrial fibrillation: insights from the apixaban for reduction in stroke and other thromboembolic events in atrial fibrillation (ARISTOTLE) trial. Circulation. 2014;130:1847-58.
19. Ridker PM. Inflammation, C-reactive protein, and cardiovascular disease: moving past the marker versus mediator debate. Circ Res. 2014;114:594-5.
20. de Lemos JA, McGuire DK, Drazner MH. B-type natriuretic peptide in cardiovascular disease. Lancet. 2003;362:316-22.
21. Netala VR, Hou T, Wang Y, Zhang Z, Teertam SK. Cardiovascular biomarkers: tools for precision diagnosis and prognosis. Int J Mol Sci. 2025;26:3218.
22. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial intelligence in cardiology. J Am Coll Cardiol. 2018;71:2668-79.
23. Ibrahim NE, Januzzi JL Jr. Established and emerging roles of biomarkers in heart failure. Circ Res. 2018;123:614-29.
24. Neufeld G, Kessler O. The semaphorins: versatile regulators of tumour progression and angiogenesis. Nat Rev Cancer. 2008;8:632-45.
25. Cook NR, Buring JE, Ridker PM. The effect of including C-reactive protein in cardiovascular risk prediction models for women. Ann Intern Med. 2006;145:21-9.
26. Zhang YJ, Sun Y, Zhao YB, Ma D. α-HBDH is a superior to LDH in predicting major adverse cardiovascular events in patients with acute aortic dissection. Heliyon. 2024;10:e29155.
27. O'Donoghue ML, Morrow DA, Cannon CP, Jarolim P, Desai NR, Sherwood MW, et al. Multimarker risk stratification in patients with acute myocardial infarction. J Am Heart Assoc. 2016;5:e002586.

28. Sinning C, Ojeda F, Wild PS, Schnabel RB, Schwarzl M, Ohdah S, et al. Midregional proadrenomedullin and growth differentiation factor-15 are not influenced by obesity in heart failure patients. Clin Res Cardiol. 2017;106:401-10.
29. Gaggin HK, Januzzi JL Jr. Biomarkers and diagnostics in heart failure. Biochim Biophys Acta. 2013;1832:2442-50.
30. Porcel JM, Bielsa S, Morales-Rull JL, Civit C, Cao G, Light RW, et al. Comparison of pleural N-terminal pro-B-type natriuretic peptide, midregion pro-atrial natriuretic peptide and mid-region pro-adrenomedullin for the diagnosis of pleural effusions associated with cardiac failure. Respirology. 2013;18:540-5.
31. Kolditz M, Halank M, Schiemanck CS, Schmeisser A, Höffken G. High diagnostic accuracy of NT-proBNP for cardiac origin of pleural effusions. Eur Respir J. 2006;28:144-50.
32. Porcel JM. Utilization of B-type natriuretic peptide and NT-proBNP in the diagnosis of pleural effusions due to heart failure. Curr Opin Pulm Med. 2011;17:215-9.
33. Takamatsu H, Kumanogoh A. Diverse roles for semaphorin-plexin signaling in the immune system. Trends Immunol. 2012;33:127-35.
34. Ibanez B, James S, Agewall S, Antunes MJ, Bucciarelli-Ducci C, Bueno H, et al. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The task force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). Eur Heart J. 2018;39:119-77.
35. Wanschel A, Seibert T, Hewing B, Ramkhelawon B, Ray TD, van Gils JM, et al. Neuroimmune guidance cue Semaphorin 3E is expressed in atherosclerotic plaques and regulates macrophage retention. Arterioscler Thromb Vasc Biol. 2013;33:886-93.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.37928
Eurasian J Emerg Med. 2026;25: 394-402

The Relationship between Optic Nerve Sheath Diameter, Optic Disc Edema, and Mortality in Patients with Septic Encephalopathy: A Prospective Observational Study

Muhammed Akay, Nazire Belgin Akıllı, Ramazan Köylü

Konya City Hospital, Clinic of Emergency Medicine, Konya, Türkiye

Abstract

Aim: Our objective was to examine ocular ultrasonography findings in sepsis-related encephalopathy and assess its relationship with mortality.

Materials and Methods: Male and female patients over the age of 18 who were diagnosed with septic encephalopathy and whose relatives or self-consent were given to participate in the study were included in the study. Healthy adult patients over 18 years of age were included in the control group. Demographic and laboratory data, glasgow coma scale, Sequential Organ Failure Assessment (SOFA), quick SOFA, APACHE II scores, ocular ultrasonographic results and death status were noted.

Results: 100 patients and 30 control groups were included in the study. While the right optic nerve sheath diameter (ONSD) values were 0.49±0.006 cm in the patient group and 0.40±0.02 cm in the control group, the left ONSD values were 0.48±0.006 cm in the patient group and 0.41±0.02 cm in the control group. While the right optic disc elevation (ODE) values were 0.009±0.03 cm in the patient group and 0.00±0 cm in the control group, the left ODE values were 0.005±0.01 cm in the patient group and 0.00±0 cm in the control group ($p<0.05$). Sensitivity for left ONSD >0.48: 72.2 (46.5-90.3), specificity: 74.3 (63.6-83.4), positive predictive value: 38.2 (28-49.7), negative predictive value: 92.4 (85.1-96.3) (area under the curve 0.75 and p: 0,0003).

Conclusion: ONSD and ODE are enlarged in patients with septic encephalopathy compared to healthy controls. Inpatients with septic encephalopathy, ONSD and ODE measurement can give preliminary information about intracranial pressure, clinical course and mortality.

Keywords: Septic encephalopathy, optic nerve sheath diameter, optic disc elevation, pocus

Introduction

Sepsis is a life-threatening clinical condition accompanied by organ dysfunction, resulting from inadequate and inappropriate host response to infection (1). It is a serious condition that affects millions of people worldwide every year and results in the death of one in four patients (2). The mortality attributed to sepsis varies based on patient-related and independent risk factors, the size of the healthcare facility, early diagnosis of the condition, and prompt initiation of treatment for patients with delayed diagnosis and treatment, even in cases of eventual recovery, a significant decline is observed in the quality of life after the disease (3,4). The development of a systemic inflammatory response in sepsis induces hemostatic changes, leading to subsequent organ dysfunction (5). One of these is sepsis-associated encephalopathy (SAE), a common but poorly understood neurological complication of sepsis. It is characterized by generalized brain dysfunction associated with an infection in another part of the body, in the absence of any sedation, analgesics, or benzodiazepines, and without an overt central nervous system infection. The pathophysiology of SAE is complex and multifactorial, involving a number of interconnected mechanisms, including vascular injury, endothelial activation, blood–brain barrier disruption, altered brain signaling, brain inflammation, and apoptosis. Clinical manifestations of SAE can range from mild symptoms, such as weakness and lack of concentration, to deep coma. The lack of any specific research or biomarkers and the widespread use of sedation in critically ill patients make the assessment of

Corresponding Author: Muhammed Akay PhD, Konya City Hospital, Clinic of Emergency Medicine, Konya, Türkiye
E-mail: makay686@gmail.com **ORCID ID:** orcid.org/0000-0002-6371-1518
Cite this article as: Akay M, Akıllı NB, Köylü R. The relationship between optic nerve sheath diameter, optic disc edema, and mortality in patients with septic encephalopathy: a prospective observational study.. Eurasian J Emerg Med. 2026;25: 394-402

Received: 12.12.2025
Accepted: 28.07.2026
Published: 04.09.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

cognitive dysfunction difficult (6). When accompanying sepsis, this condition, known as septic encephalopathy, can increase mortality rates from 26% to 49% (7). It is important to recognize SAE because of its high mortality rate. SAE can manifest in the initial stages of sepsis, even preceding the fulfillment of diagnostic criteria for sepsis (6).

Increased intracranial pressure (ICP) is a condition that arises from cerebral edema and can result from either traumatic or metabolic factors. Cerebral edema and increased ICP may occur during the course of sepsis. Intracranial catheters are used as an invasive method for measuring ICP, whereas computed tomography, transcranial Doppler sonography, ophthalmoscopy, and ultrasonography (USG) measurement of optic nerve sheath diameter (ONSD) serve as noninvasive techniques for ICP measurement. These diagnostic methods have both advantages and disadvantages when compared to each other (8). Studies indicate that USG measurement of ONSD demonstrates high sensitivity and specificity in detecting increased ICP (9). Compared to other methods, USG assessment of ICP is faster, simpler, reproducible, noninvasive, and can be conducted at the bedside.

In this study, our objective was to examine ocular USG findings in sepsis-related encephalopathy and assess its relationship with mortality.

Materials and Methods

This study was conducted as a prospective cohort study at the Emergency Department of Konya City Hospital from October 2021 to June 2022. Ethics committee approval for the study was obtained from the Necmettin Erbakan University Non-Drug and Non-Medical Device Research Ethics Committee (decision no: 2021/3413(7074), date: 17.09.2021) conducted in accordance with the principles outlined in the Declaration of Helsinki. The study included 100 men and women patients aged over 18 diagnosed with septic encephalopathy, all of whom provided consent either personally or through their relatives. A control group comprising 30 healthy adults aged over 18 was included in the study. Electronic and written documentation were used to record data from both the groups.

No formal a priori sample size calculation was performed before patient enrollment. After completion of the study, a sample size adequacy assessment was conducted for the primary comparison of ONSD between patients with SAE and healthy controls. Assuming a two-sided alpha level of 0.05, 80% power, an allocation ratio of approximately 3:1, and an anticipated moderate-to-large standardized effect size of 0.60 for the difference in ONSD, at least 89 patients and 30 controls were required. The final study population of 100 patients and 30 controls was therefore considered adequate for the primary between-group comparison. Analyses comparing survivors and non-survivors were considered exploratory because of the relatively small number of non-survivors.

We excluded patients under 18 years old, those with traumatic injuries, anatomical eye defects, optic neuritis, as well as individuals with an unassessable glasgow coma scale (GCS) score due to Alzheimer's disease and dementia. Patients with structural central nervous system disorders, such as intracranial mass, hemorrhage, or infarction, those with sepsis due to meningitis or encephalitis, acute renal failure, hyponatremia, hypernatremia, diabetic ketoacidosis, hyperosmolar nonketotic coma, hypoglycemia, Wernicke's encephalopathy, toxic drug overdose, hepatic encephalopathy, or suspected encephalopathy from metabolic or toxic abnormalities, pregnant patients, and those who declined to provide consent were also excluded.

Sepsis was diagnosed based on the current guideline criteria, identifying it as an inappropriate host response to the existing infection, subsequently causing organ dysfunction (10). The diagnosis of infection was confirmed by culture results.

Septic encephalopathy was defined as a GCS score below 15 or the presence of delirium symptoms (inattention, disorientation, altered thinking, decreased psychomotor activity, and/or agitation), accompanied by cognitive and neuropsychiatric impairment identified by healthcare personnel (physicians, nurses) or the relatives of a patient in a sepsis clinic (11).

Age, gender, systolic and diastolic blood pressures, pulse rate, saturation levels, GCS scores, Sequential Organ Failure Assessment (SOFA) scores, quick SOFA (qSOFA) scores, APACHE II scores, hematologic and biochemical markers, site and source of infection, ocular USG findings, duration of hospitalization, and in-hospital mortality were recorded.

Optic USG was performed after the patient was diagnosed with septic encephalopathy in the emergency department. The practitioner was trained and experienced in ocular USG and had 4 years of experience in ocular USG. All examinations were performed using a Mindray DC-80 ultrasound system with a linear probe (5-12 MHz). The procedure was conducted in a blinded manner, ensuring that participant information (patient and control) remained undisclosed. During the examination, patients were in the supine position with the head in a neutral position. Ocular USG was used to measure both the ONSD and the optic disc height (Figure 1 and 2). Using a closed-eye technique, a linear probe was placed directly onto the eyelid where gel was applied, and the eyeball was visualized in the transverse plane. ONSD was measured transversely, positioned

0.3 cm behind the papilla in the eyeball. To measure optic disc height, ultrasound imaging was used to determine the distance between the anterior apex of the optic disc and its intersection with the posterior surface of the globe. The measurements were performed three times for each eye, and the average value was calculated. The ocular USG findings were compared between the groups with septic encephalopathy and the healthy controls. Based on the acquired findings, the association between ocular USG findings and mortality was investigated.

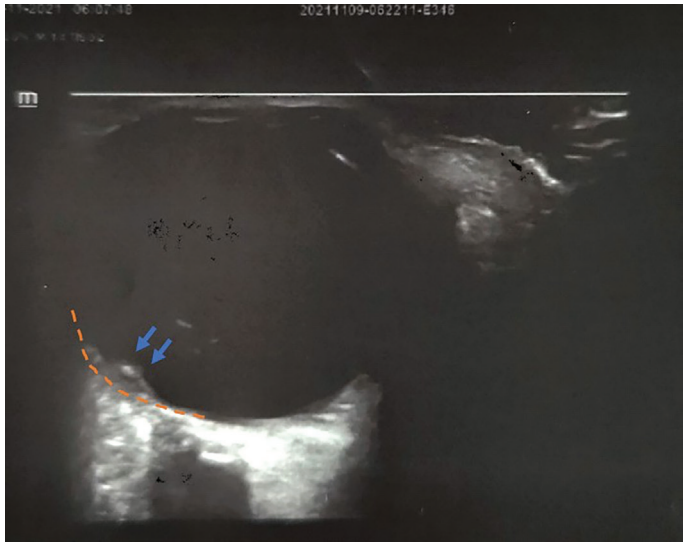


Figure 1. Ultrasonographic appearance of optic disc edema. Representative ocular ultrasonography image demonstrating optic disc elevation consistent with optic disc edema. The area protruding from the optical disk boundary, indicated by two blue arrows and orange dashed lines, is the optical disk elevation

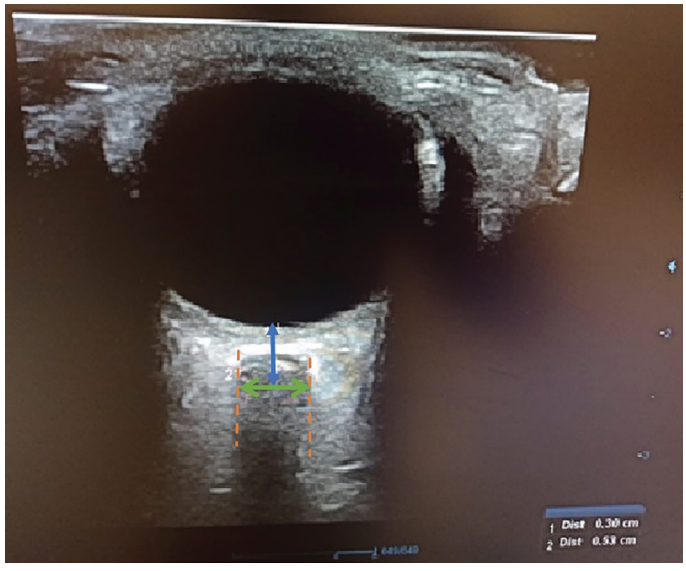


Figure 2. Ultrasonographic measurement of optic nerve sheath diameter. Representative ocular ultrasonography image showing the measurement of optic nerve sheath diameter, performed 3 mm posterior to the optic disc. The blue line indicates a distance of 3 mm from the optic disc. The dashed orange line indicates the boundaries of the optic nerve. The horizontal green line indicates the diameter of the optic nerve sheath

The data were entered into the SPSS 20 database for statistical analysis. The conformity of the variables to normal distribution was examined both visually (histogram and probability graphs) and analytically (Kolmogorov-Smirnov and Shapiro-Wilk tests). Descriptive statistics were reported as mean ± standard deviation (SD) for normally distributed variables, median [interquartile range (IQR)] for non-normally distributed variables, and n (%) for categorical variables. The Student's t-test was applied for normally distributed variables, the Mann-Whitney U test for non-normally distributed variables, and the Fisher's exact test (instead of the chi-square test) for categorical variables in comparisons between two independent groups. ROC curve analysis was performed to evaluate the discriminatory performance of ONSD measurements for mortality. The area under the curve (AUC) with 95% confidence intervals was calculated. The optimal cut-off values were determined using the Youden index method, and sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and likelihood ratios were calculated. A p value <0.05 was considered statistically significant.

Results

When analyzing the demographic characteristics of the study groups, the mean age was 70.07±15.9 years in the patient group and 69.2±12.7 years in the control group. There was no statistically significant difference between the groups in terms of age (p>0.05). In the patient group, 44% (n=44) were women, whereas 56% (n=56) were men. In the control group, the gender distribution was 50% women (n=15) and 50% men (n=15). When comparing the patient and control groups based on gender, no significant difference was observed (p>0.05). Demographic characteristics are presented in Table 1.

In our study, the source of infection was unidentified in 42 (42%) cases of pneumonia, 43 (43%) cases of urinary tract infection, 3 (3%) cases of acute gastroenteritis, and 5 (5%) cases.

Of the patients with infection, the source of infection was not found in 66 (66%). *E. coli* was the source in 13 (13%) patients, *Klebsiella pneumonia* in 5 (5%) patients, and Gram-positive *cocci* in 3 (3%) patients. The cause of infection was unidentified in 66 (66%) cases. Sources and causes of infection are shown in Table 2.

Out of the 100 patients included in the study, 18 did not survive, while 82 were discharged with recovery.

The mean right ONSD was 0.49±0.006 cm in the patient group and 0.40±0.02 cm in the control group. The mean left ONSD was

Table 1. Demographic characteristics of study population				
Feature	**Subgroup**	**Case group (n:100)**	**Control group (n:30)**	**p value**
Age (year) mean ± SD		70.07±15.9	69.2±12.7	0.77
Gender %(n)	Male	56% (n=56)	50% (n=15)	0.56
	Woman	44% (n=44)	50% (n=15)	
Vital signs	SBP mmHg mean ± SD	82.45±0.87		
	DBP mmHg mean ± SD	49.9±0.93		
	Pulse/min. mean ± SD	97.56±1.32		
Laboratory	WBC 0^3/µL median (IQR)	15.05 (9.43)		
	HGB g/Dl median (IQR)	12.7 (2.8)		
	PLT 10^3/µL median (IQR)	239.5 (156.25)		
	Glucose mg/dL median (IQR)	133 (68.5)		
	BUN mg/dL median (IQR)	22 (12.5)		
	Creatine mg/dL median (IQR)	1.2 (0.58)		
	Sodium mmol/L mean ± SD	136.46±0.58		
	Potasıum mmol/L mean ± SD	4.46±0.065		
	CRP mg/L mean ± SD	223.29±9.49		
	Procalsitonin µg/L median (IQR)	1.62 (9.42)		
Scoring	GCS median (IQR)	13 (2)		
	qSOFA median (IQR)	2 (1)		
	SOFA median (IQR)	6 (2)		
	APACHE median (IQR)	14 (10)		
Losıng period (day) median (IQR)		9.5 (8)		
ONSD	ONSD right (cm) mean ± SD	0.49±0,006	0.40±0,02	0.00
	ONSD left (cm) mean ± SD	0.48±0.006	0.41±0.02	0.00
ODE	ODE right (cm) mean ± SD	0.009 ±0.03	0.00±0	0.01
	ODE left (cm) mean ± SD	0.005±0.01	0.00±0	0.006
Mortality %(n)		18% (n=18)		

SD: Standard deviation, IQR: Interquartile range, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, GCS: Glasgow coma score, WBC: White blood cell, HBG: Hemoglobin, PLT: Platelet, BUN: Blood urea nitrogen, CRP: C-reactive protein, USG: Ultrasonography, qSOFA: quick Sepsis Related Organ Failure Assessment, SOFA: Sequential Organ Failure Assessment, ONSD: Optic nerve sheath diameter, ODE: Optic disc elevation

Table 2. Causes of infection sources	
Escherichia coli	**13**
Enterococcus	2
Klebsiella pneumoniae	5
Staphylococcus aureus	1
Staphylococcus capitis	1
Leucocytes in sputum	2
Acinetobacter baumannii	1
Gram-positive cocci	3
Enterococcus faecium	1
Morganella morganii	1
Coagulase-negative *Staphylococcus*	1
Group A beta-hemolytic *Streptococcus*	2
Gram-negative bacilli	1
Undiagnosed	66

0.48±0.006 cm in the patients group and 0.41±0.02 cm in the control group. There was a significant difference between the patient and control groups in both right and left ONSD values (p<0.05). Among the other ocular USG parameters, the mean right optic disc elevation (ODE) measured 0.009±0.03 cm in the patient group and 0.00±0 cm in the control group. The mean left ODE was 0.05±0.1 cm in the patient group and 0.00±0 cm in the control group. Statistically significant differences were observed in both right and left ODE between the case and control groups (p<0.05).

In the comparison of survivors and non-survivors, the median (IQR) age of the 18 non-survivors was 77 (14) years, whereas the median (IQR) age of the 82 survivors was 73 (16) years (p=0.17). Among the non-survivors, 66.7% (12/18) were male, compared with 53.6% (44/82) of survivors (p=0.31). The median right ONSD was 0.52 (0.12) cm in non-survivors and 0.46 (0.06) cm in survivors, showing a difference close to statistical significance (p=0.06). The median left ONSD was significantly higher in non-survivors than in survivors [0.52 (0.11) cm vs. 0.46 (0.06) cm, respectively; p=0.01]. No significant differences were observed in right-eye ODE [0 (0.025) vs. 0 (0), p=0.08] or left-eye ODE [0 (0) vs. 0 (0), p=0.95]. A comparison of vital signs and laboratory parameters between survivors and non-survivors is presented in Table 3.

Sensitivity for right ONSD > 0.5 cm: 55.5 (30.8-78.5), specificity: 81.7 (71.6-89.4), PPV: 40 (26.5-55.3), NPV: 83.3 (83.2-93.4). (AUC is 0.70 and p: 0.003).

Sensitivity for left ONSD > 0.48 cm: 72.2 (46.5-90.3), specificity: 74.3 (63.6-83.4), PPV: 38.2 (28-49.7), NPV: 92.4 (85.1-96.3). (AUC 0.75 and p: 0.0003) (Figure 3). Table 4 presents the performance of ONSD in predicting mortality.

Discussion

Our study is a prospective observational study investigating the diagnostic and prognostic significance of ONSD and ODE in

Table 3. Comparison of dead and living patients

	Death n=18	Live n=82	p
Age year median (IQR)	77 (14)	73 (16)	0.17
Gender n (%)	%66.7 (12)	%53.6 (44)	0.31
Pulse min (%)	115 (16.3)	90 (10)	0.00
SO_2 %median (IQR)	87.5 (6)	90 (5)	0.00
SBP mmHG median (IQR)	80 (10)	90 (10)	**0.00**
DBP MmHG median n (IQR)	40 (10)	50 (20)	**0.001**
$WBC0^3/\mu L$ median (IQR)	12.6 (10.5)	15.65 (9.8)	0.64
HGB g/Dl median (IQR)	10.9 (4.3)	12.7 (2.3)	0.038
PLT $10^3/\mu L$ median (IQR)	99.5 (224)	250 (145.8)	**0.01**
Glucose mg/Dl median (IQR)	141.5 (52.5)	131.5 (86)	0.8
BUN mg/Dl median (IQR)	21.5 (10.7)	23.5 (13.25)	0.78
Creatine mg/Dl median (IQR)	1.3 (0.6)	1.1 (0.5)	0.19
Sodium mmol/ median (IQR)	135 (9)	137 (6)	0.35
Potasium mmol/L median (IQR)	4.5 (0.9)	4.4 (0.8)	0.52
CRP mg/L median (IQR)	205 (195)	232 (140.8)	0.55
Procalsitonin µg/L median (IQR)	1.59 (50.3)	1.67 (8)	0.28
APACHE median (IQR)	27 (12)	12 (7)	**0.00**
GCS median (IQR)	12 (3)	13 (1)	**0.00**
qSOFA median (IQR)	3 (0)	2 (0)	**0.00**
SOFA median (IQR)	12.5 (3)	5 (3)	**0.00**
ONSD rıght (cm) median (IQR)	0.52 (0.12)	0.46 (0.006)	0.06
ONSF left (cm) median (IQR)	0.52 (0.11)	0.46 (0.006)	**0.01**
ODE right (cm) median (IQR)	0 (0.025)	0 (0)	0.08
ODE left (cm) median (IQR)	0 (0)	0 (0)	0.95

SD: Standard deviation, IQR: Interquartile range, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, GCS: Glasgow coma score, WBC: White blood cell, HBG: Hemoglobin, PLT: Platelet, BUN: Blood urea nitrogen, CRP: C-reactive protein, USG: Ultrasonography, qSOFA: quick Sepsis Related Organ Failure Assessment, SOFA: Sequential Organ Failure Assessment, ONSD: Optic nerve sheath diameter, ODE: Optic disc elevation

patients admitted to the emergency department with sepsis-related encephalopathy. Based on our study findings, patients diagnosed with septic encephalopathy exhibited enlarged ONSD and ODE in comparison to the healthy control group. Among patients diagnosed with SAE, both eyes showed higher ONSD measurements in non-survivors when compared to survivors. Although not statistically significant, mortality was numerically higher among patients with SAE who developed ODE compared with those without ODE (35% vs. 15%, p=0.06).

In sepsis, almost every organ can be affected or fail. In the respiratory system, sepsis can induce acute respiratory distress syndrome, whereas in the gastrointestinal tract, the effects can range from edema and inflammation on the bowel wall to ulcer, ileus, and hemorrhage. Hypovolemia, hypotension, and renal vasoconstriction can cause acute tubular necrosis and acute kidney failure. SAE is a severe neurological syndrome characterized by extensive brain dysfunction resulting from the body's abnormal response to infection (12,13).

Septic encephalopathy is one of the main manifestations of organ dysfunction caused by sepsis, excluding clinical or laboratory evidence of central nervous system infection, structural abnormality, or other forms of encephalopathy, such as hepatic encephalopathy or uremic encephalopathy. Septic encephalopathy refers to extensive brain dysfunction caused by sepsis and mainly manifests as delirium, cognitive impairment, reduced learning and memory abilities, and coma (14,15). Patients with septic encephalopathy may have any degree of disorientation and cognitive impairment, ranging from mild disturbance of consciousness to deep coma (12,16). In delirium due to sepsis, there is hypoactivity rather than hyperactivity (17).

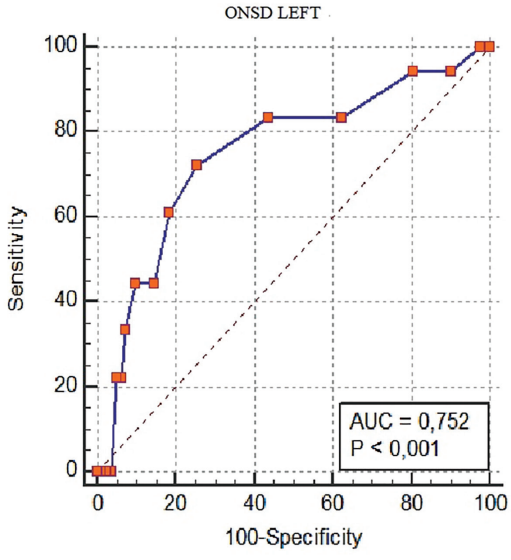


Figure 3. ROC curve analysis of left optic nerve sheath diameter (ONSD) for mortality. ROC curve demonstrating the discriminatory performance of left ONSD for mortality among patients with septic encephalopathy

AUC: Area under the curve

The underlying mechanism might be cerebral microvascular cell dysfunction, compromised integrity of the blood–brain barrier, mitochondrial malfunction, activation of microglia and astrocytes, and subsequent neuronal death (18,19). These mechanisms are not uniform throughout the brain, leading to varied clinical presentations based on the affected areas (20,21). No specific biomarker or radiological imaging method exists for the definitive diagnosis of septic encephalopathy. There are several biomarkers that indicate brain damage (22-24). These biomarkers have low sensitivity and specificity and have little clinical relevance (25). The use of blood biomarkers for diagnosing SAE faces challenges due to undefined optimal assessment timing, uncertainty about the exact onset of SAE, inconclusive findings on prognostic utility and precise thresholds, and limited accessibility to specific biomarkers, hindering their routine clinical application (26). In 20% of patients with septic encephalopathy, electroencephalography (EEG) reveals changes similar to those observed in nonconvulsive status epilepticus (27). At the same time, different wave activities can be monitored. However, sedative drugs, such as benzodiazepines, barbiturates and propofol, may cause similar abnormalities in EEG. Furthermore, EEG lacks specificity in assessing SAE, as similar abnormalities might manifest in various forms of encephalopathy (26). Radiological imaging studies rarely reveal lesions. In 30% of patients with septic shock, magnetic resonance imaging scans might display hyperintensity in the white matter, a finding suggestive of ischemia (28,29). While there have been multiple studies on the use of evoked potentials, their application is constrained by several factors, limited availability of devices in the intensive care setting, challenges in interpretation, and their moderate prognostic value within the sepsis population (26).

Table 4. Performance characteristics of ONSD in predicting mortality

	Sensitivity (%95 CI)	Specificity (%95 CI)	+LR (%95 CI)	-LR (%95 CI)	+PV (%95 CI)	-PV (%95 CI)	AUC (%95 CI)	p
Right ONSD >0.5	55.5 (30.8-78.5)	81.7 (71.6-89.4)	3.04 (1.6-5.6	0.54 (0.3-0.9)	40 (26.5-55.3)	83.3 (83.2-93.4)	0.70 (0.58-0.79)	0.003
Left ONSD >0.48	72.2 (46.5-90.3)	74.3 (63.6-83.4)	2.82 (1.8-4.5)	0.37 (0.2-0.8)	38.2 (28-49.7)	92.4 (85.1-96.3)	0.75 (0.69-0.86)	0.0003

ONSD: Optic nerve sheath diameter, CI: Confidence İnterval, LR: Likelihood Ratio, PV: Predictive Value, AUC: Area under the curve

The diagnostic criteria and potential risk factors for septic encephalopathy remain poorly understood, lacking a reliable tool for the clinical assessment of neurological dysfunction associated with sepsis (30). As in many other organ dysfunctions, sepsis induces cerebral damage, resulting in cerebral edema. Cerebral edema secondary to SAE is recognized as one of the leading causes of complications and death in patients with sepsis, and early detection of intracranial hypertension is of great importance for timely intervention and improved prognosis (31). There are invasive and noninvasive diagnostic methods to detect brain edema.

Ocular USG, an inexpensive and easily conducted noninvasive bedside technique, has previously demonstrated its utility as an advantageous tool for clinicians in evaluating disease severity and clinical progression. In the initial studies conducted by Hansen et al. (32) in 1994 and later studies by Ohle (9) an association was established between the increase in ONSD and increased ICP, demonstrating high sensitivity and specificity. In a study conducting ultrasonographic evaluation of the optic nerve in patients with severe traumatic brain injury, those with ICP exceeding 22 mmHg exhibited significantly higher ONSD and ODE values. It was found that an ONSD greater than 0.72 demonstrated a sensitivity of 82% (48%-98%) and a specificity of 79% (70%-86%). An ODE greater than 0.04 cm demonstrated a sensitivity of 90% (56%-100%) and a specificity of 71% (61%-79%) (33). In another study in which USG ODE evaluation was performed by emergency physicians, maximum disc height greater than 0.6 cm predicted the presence of ODE noted on fundoscopic examination with a sensitivity of 82% (48%-98%) and a specificity of 76% (50%-93%) (34).

Only one study was identified that examined the relationship between SAE and ocular USG. In the study conducted by Yang et al. (31), ultrasound examination of the ONSD was conducted on 142 patients with sepsis. These patients were categorized into three groups: those who did not develop SAE, those who developed SAE, and those who developed SAE but eventually recovered. The median ONSD was 0.51 (0.475-0.54) cm, 0.59 (0.56-0.625) cm, respectively) and 0.535 (0.5075-0.55) cm across the respective patient groups. In their study, measurements were obtained with the patient's head elevated at a 30-degree angle. Patients with SAE had a significantly wider ONSD than patients without SAE ($p<0.001$) and patients with recovered SAE. Their study did not examine the association of ODE and ocular USG findings with mortality.

In our study, upon comparing the patient group with the control group, the mean right ONSD was 0.49 ± 0.006 in the patient group and 0.40 ± 0.02 in the control group, while the mean left ONSD measured 0.48 ± 0.006 in the patient group and 0.41 ± 0.02 in the control group. Additionally, 16% of the patients exhibited ODE. Among the patients with ODE, 35% did not survive, whereas 15% of those without USG ODE died ($p=0.06$). Hence, it is plausible that the ONSD might be enlarged in patients with sepsis-related encephalopathy. Our findings suggest that optic USG may represent a promising bedside tool for the assessment of patients with septic encephalopathy, although further validation studies are required. The increase in ONSD and the presence of ODE may be associated with adverse outcomes; however, further studies with larger sample sizes are needed to clarify their prognostic value. Specifically, the left ONSD demonstrates acceptable predictive capability for mortality, with an AUC of 0.75. It may also be useful in detecting low-risk patients with a NPV of 92.4%. Regarding cerebral resuscitation in patients exhibiting enlarged ONSD or ODE, directing interventions toward preventing fever, regulating blood pressure and ventilation, maintaining euglycemia, preventing electrolyte imbalances, and administering anti-edema therapy could potentially enhance mortality outcomes. In fact, Sonneville et al. (26) highlights the importance of preventing secondary damage in SAE and emphasize that precautions should be taken just like post-resuscitation care and neurocritical care guidelines, although they are not mentioned in sepsis guidelines. Laboratory tests including EEG and sensory evoked potentials used in the diagnosis of septic encephalopathy are both more expensive and not always easy to perform under intensive care conditions. Interpretation also requires expertise. USG represents an accessible, inexpensive diagnostic and monitoring modality, readily applicable by clinicians at the bedside. Its ease of training and widespread utilization in daily practice by intensive care and emergency medicine specialists underscore its significance. Considering various factors contributing to encephalopathy, including the use of sedatives and neurotoxic antibiotics, particularly during the intensive care unit management of sepsis patients, neurological assessments can be complex. We believe that patient monitoring through USG could serve as a valuable aid for clinicians in navigating these challenges.

Study Limitations

Our study is one of the first studies investigating ocular USG findings in patients with septic encephalopathy. However, it also has some limitations. Firstly, this is the single-center design of our study and small sample size. Secondly, the diagnosis of septic encephalopathy relied on an exclusionary diagnostic approach, considering patient complaints, clinical findings, and investigations, a method commonly employed in similar studies. EEG and sensory evoked potentials were not performed because of the cost and the fact that these examinations can be performed in the electrophysiology laboratory of our hospital. A third limitation involves the

selection of the control group from healthy volunteers. In fact, it would have been better if the control group consisted of patients with sepsis without encephalopathy. However, studies have shown that EEG can reveal brain abnormalities in 50% of patients with laboratory evidence of bacteremia but no evidence of clinical encephalopathy on physical examination (with relatively preserved cognition) (35). Brain dysfunction is one of the earliest forms of organ dysfunction in sepsis, and sepsis-induced delirium occurs in as high as 70% of patients (36). The encephalopathy clinic may also occur later in the course of sepsis. The fact that we used qSOFA in the inclusion criteria according to the guidelines at that time and that this score includes those with low GCS made it difficult for us to take the control group from patients with sepsis without encephalopathy. Recognizing the potential bias arising from these factors, we composed the control group using healthy individuals. Multicenter and more comprehensive studies are needed. Another limitation of our study is that all ocular USG measurements were performed by a single experienced operator. Therefore, inter-rater reliability could not be assessed, and intra-rater reproducibility was not evaluated with repeated measurements. Future studies including multiple trained operators and repeated measurements are needed to further assess the reproducibility and clinical applicability of ocular USG-based ONSD measurements.

Conclusion

In patients with SAE, the ONSD is wider than in healthy individuals and ODE may develop. An expanded ONSD can be used to predict mortality with an AUC of 0.75. There is currently no standard biochemical, radiological or electrophysiologic test for the diagnosis of SAE. Studies with USG are needed in this field and optical USG is promising in this respect. Moreover, we believe that further studies are needed to investigate the effect of interventions aiming to improve cerebral resuscitation on sepsis-related mortality in patients with enlarged ONSD and ODE.

Ethics

Ethics Committee Approval: Ethics committee approval for the study was obtained from the Necmettin Erbakan University Non-Drug and Non-Medical Device Research Ethics Committee (decision no: 2021/3413(7074), date: 17.09.2021).

Informed Consent: This study was conducted as a prospective cohort study.

Data availability: The collected data sets are available from the corresponding author upon reasonable request in anonymized form.

Acknowledgments: The authors thank Enago - https://www.enago.com.tr/ceviri/ for their assistance in manuscript translation and editing.

Footnotes

Authorship Contributions

Surgical and Medical Practices: M.A., N.B.A., R.K., Concept: M.A., R.K., Design: M.A., N.B.A., R.K., Data Collection or Processing: M.A., N.B.A., Analysis or Interpretation: M.A., R.K., Literature Search: M.A., N.B.A., R.K., Writing: M.A., N.B.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA. 2016;315:801-10.
2. Mayr FB, Yende S, Angus DC. Epidemiology of severe sepsis. Virulence. 2014;5:4-11.
3. Martin GS, Mannino DM, Eaton S, Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. N Engl J Med. 2003 Apr 17;348:1546-54.
4. Baykara N, Akalın H, Arslantaş MK, Hancı V, Çağlayan Ç, Kahveci F, et al. Epidemiology of sepsis in intensive care units in Turkey: a multicenter, point-prevalence study. Crit Care. 2018;22:93.
5. Hotchkiss RS, Karl IE. The pathophysiology and treatment of sepsis. N Engl J Med. 2003;348:138-50.
6. Chaudhry N, Duggal AK. Sepsis associated encephalopathy. Adv Med. 2014;2014:762320. Epub 2014 Sep 30.
7. Green R, Scott LK, Minagar A, Conrad S. Sepsis associated encephalopathy (SAE): a review. Front Biosci. 2004;9:1637-41.
8. Soldatos T, Karakitsos D, Chatzimichail K, Papathanasiou M, Gouliamos A, Karabinis A. Optic nerve sonography in the diagnostic evaluation of adult brain injury. Crit Care. 2008;12:R67.
9. Ohle R, McIsaac SM, Woo MY, Perry JJ. Sonography of the optic nerve sheath diameter for detection of raised intracranial pressure compared to computed tomography: a systematic review and meta-analysis. J Ultrasound Med. 2015;34:1285-94.
10. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. Intensive Care Med. 2017;43:304-77. Epub 2017 Jan 18.
11. Chen J, Shi X, Diao M, Jin G, Zhu Y, Hu W, et al. A retrospective study of sepsis-associated encephalopathy: epidemiology, clinical features and adverse outcomes. BMC Emerg Med. 2020;20:77.
12. Fu Q, Wu J, Zhou XY, Ji MH, Mao QH, Li Q, et al. NLRP3/Caspase-1 Pathway-induced pyroptosis mediated cognitive deficits in a mouse model of sepsis-associated encephalopathy. Inflammation. 2019;42:306-18.
13. Bouadma L, Sonneville R, Garrouste-Orgeas M, Darmon M, Souweine B, Voiriot G, et al. Ventilator-associated events: prevalence, outcome, and relationship with ventilator-associated pneumonia. Crit Care Med. 2015;43:1798-806.

14. Gofton TE, Young GB. Sepsis-associated encephalopathy. Nat Rev Neurol. 2012;8:557-66. Epub 2012 Sep 18.
15. Zampieri FG, Park M, Machado FS, Azevedo LC. Sepsis-associated encephalopathy: not just delirium. Clinics (Sao Paulo). 2011;66:1825-31.
16. Sonneville R, de Montmollin E, Poujade J, Garrouste-Orgeas M, Souweine B, Darmon M, et al. Potentially modifiable factors contributing to sepsis-associated encephalopathy. Intensive Care Med. 2017;43:1075-84.
17. Polito A, Eischwald F, Maho AL, Polito A, Azabou E, Annane D, et al. Pattern of brain injury in the acute setting of human septic shock. Crit Care. 2013;17:R204.
18. Huang M, Wei R, Wang Y, Su T, Li P, Chen X. The uremic toxin hippurate promotes endothelial dysfunction via the activation of Drp1-mediated mitochondrial fission. Redox Biol. 2018;16:303-13.
19. Ren C, Yao RQ, Zhang H, Feng YW, Yao YM. Sepsis-associated encephalopathy: a vicious cycle of immunosuppression. J Neuroinflammation. 2020 Jan 10;17:14.
20. Mazeraud A, Pascal Q, Verdonk F, Heming N, Chrétien F, Sharshar T. Neuroanatomy and physiology of brain dysfunction in sepsis. Clin Chest Med. 2016;37:333-45.
21. Cunningham C, Maclullich AM. At the extreme end of the psychoneuroimmunological spectrum: delirium as a maladaptive sickness behaviour response. Brain Behav Immun. 2013;28:1-13.
22. Ehler J, Barrett LK, Taylor V, Groves M, Scaravilli F, Wittstock M, et al. Translational evidence for two distinct patterns of neuroaxonal injury in sepsis: a longitudinal, prospective translational study. Crit Care. 2017;21:262.
23. Tauber SC, Djukic M, Gossner J, Eiffert H, Brück W, Nau R. Sepsis-associated encephalopathy and septic encephalitis: an update. Expert Rev Anti Infect Ther. 2021;19:215-31.
24. Ehler J, Saller T, Wittstock M, Rommer PS, Chappell D, Zwissler B, et al. Diagnostic value of NT-proCNP compared to NSE and S100B in cerebrospinal fluid and plasma of patients with sepsis-associated encephalopathy. Neurosci Lett. 2019;692:167-73. Epub 2018 Nov 10.
25. Mazeraud A, Righy C, Bouchereau E, Benghanem S, Bozza FA, Sharshar T. Septic-associated encephalopathy: a comprehensive review. Neurotherapeutics. 2020;17:392-403.
26. Sonneville R, Benghanem S, Jeantin L, de Montmollin E, Doman M, Gaudemer A, Thy M, Timsit JF. The spectrum of sepsis-associated encephalopathy: a clinical perspective. Crit Care. 2023;27:386.
27. Hosokawa K, Gaspard N, Su F, Oddo M, Vincent JL, Taccone FS. Clinical neurophysiological assessment of sepsis-associated brain dysfunction: a systematic review. Crit Care. 2014;18:674.
28. Oddo M, Carrera E, Claassen J, Mayer SA, Hirsch LJ. Continuous electroencephalography in the medical intensive care unit. Crit Care Med. 2009;37:2051-6.
29. Finelli PF, Uphoff DF. Magnetic resonance imaging abnormalities with septic encephalopathy. J Neurol Neurosurg Psychiatry. 2004;75:1189-91.
30. Sonneville R, Verdonk F, Rauturier C, Klein IF, Wolff M, Annane D, et al. Understanding brain dysfunction in sepsis. Ann Intensive Care. 2013;3:15.
31. Yang Z, Qin C, Zhang S, Liu S, Sun T. Bedside ultrasound measurement of optic nerve sheath diameter in patients with sepsis: a prospective observational study. Crit Care. 2020;24:235.
32. Hansen HC, Helmke K, Kunze K. Optic nerve sheath enlargement in acute intracranial hypertension. Neuroophthalmology. 1994;14:345-54.
33. Agrawal D, Raghavendran K, Zhao L, Rajajee V. A prospective study of optic nerve ultrasound for the detection of elevated intracranial pressure in severe traumatic brain injury. Crit Care Med. 2020;48:e1278-85.
34. Teismann N, Lenaghan P, Nolan R, Stein J, Green A. Point-of-care ocular ultrasound to detect optic disc swelling. Acad Emerg Med. 2013;20:920-5.
35. Young GB, Bolton CF, Archibald YM, Austin TW, Wells GA. The electroencephalogram in sepsis-associated encephalopathy. J Clin Neurophysiol. 1992;9:145-52.
36. Donnelly J, Budohoski KP, Smielewski P, Czosnyka M. Regulation of the cerebral circulation: bedside assessment and clinical implications. Crit Care. 2016;20:129.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.08455
Eurasian J Emerg Med. 2026;25: 403-10

Frequent Users of a Provincial 112 Emergency Medical Service in Türkiye: A Retrospective Cross-sectional Study

Sefer Özkaya, Mehmet Serkan Yurdakul

Karaman Provincial Health Directorate, Karaman, Türkiye

Abstract

Aim: A small minority disproportionately burdens emergency medical services (EMS), yet frequent users of Türkiye's 112 system remain uncharacterised. We determined the proportion, predictors, and subgroups associated with frequent EMS use.

Materials and Methods: This retrospective cross-sectional study analysed all 112 EMS calls from Karaman Province, Türkiye, in 2025. After excluding records lacking a national identification number and records of inter-facility transfers, 26,347 calls from 18,653 patients remained. Frequent users were defined a priori as ≥5 calls/year. Patient-level multivariable logistic regression identified independent predictors; a patient-clustered call-level model separated per-encounter signal from cumulative signal.

Results: Of 18,653 patients, 430 (2.3%) were frequent users, generating 3,260 calls (12.4%). The two largest patient-level associations differed: a yellow or red triage code [adjusted odds ratio (AOR): 12.66, 95% confidence interval (CI): 7.30-21.96] carried no per-encounter signal [call-level odds ratio (OR): 0.98], whereas 24-hour recontact (AOR: 12.15, 95% CI: 9.46-15.60) retained the strongest per-encounter association (call-level OR: 4.59). The strongest clinical predictors were primary psychiatric or substance-use diagnoses (AOR: 7.99) and respiratory diagnoses (AOR: 6.53). Psychiatric diagnosis showed no association at the call level (8.1% vs 7.6%), but was present in 26.7% of frequent users versus 8.9% of others at the patient level. Estimates were stable across thresholds. Two subpopulations differed in composition: the younger characterized by psychiatric and social vulnerabilities, and the older by chronic respiratory disease; yet the strength of predictors did not differ by age.

Conclusion: Frequent users formed a small group comprising two distinct subpopulations and characterized by identifiable predictors. Coordinated primary care, mental health, and chronic disease pathways warrant evaluation; multi-centre replication is needed before policy change.

Keywords: Emergency medical services, health services needs and demand, mental health, primary health care, Türkiye

Introduction

Demand for emergency medical services (EMS) has risen across health systems over the past two decades, driven by population ageing, chronic disease and constrained primary care access (1,2). This growth is not uniform: in most systems, a small minority accounts for a disproportionate share of contacts which is increasingly read as a signal of unmet need that surfaces only when patients reach the ambulance. Türkiye's 112 EMS system, coordinated under the Ministry of Health, operates in every province through a command centre under physician oversight, yet Turkish evidence comes largely from metropolitan systems such as İzmir (2,3) and Konya (4), with little from smaller Anatolian provinces.

Frequent use is as much a public health observation as an operational one. In a systematic review, Scott et al. (5) found that 0.2-23% of EMS patients met various definitions of frequent use, generating 1.4-40% of activity. Such patients present more often with chronic conditions than with acute trauma, particularly respiratory and psychiatric diagnoses and substance-use disorders (1,6); these point upstream to unmet primary care and mental health needs, while repeated use strains resources (7,8).

No standard definition exists. In a scoping review of 19 studies, Harring et al. (9) found thresholds for frequent callers averaged 42.5 calls/year and those for frequent users 4.7 responses/year; the terms capture different populations, frequent users being the subset whose contacts result in resource deployment (9,10).

Corresponding Author: Sefer Özkaya MD, Karaman Provincial Health Directorate, Karaman, Türkiye
E-mail: seferozkaya@gmail.com **ORCID ID:** orcid.org/0000-0001-9794-7775
Cite this article as: Özkaya S, Yurdakul MS. Frequent users of a provincial 112 emergency medical service in Türkiye: a retrospective cross-sectional study. Eurasian J Emerg Med. 2026;25: 403-10

Received: 03.07.2026
Accepted: 12.08.2026
Published: 04.09.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Kjærgaard et al. (11) compared a literature-based threshold (≥43 calls/year) with percentile cut-offs in a Danish dataset, where the top 1% of callers corresponded to roughly five or more calls per year. We adopted ≥5 calls/year a priori, as this fixed count identifies a clinically meaningful group and, unlike a size-dependent percentile, stays comparable across systems.

The evidence originates almost exclusively from Northern and Western Europe (1,7,8,10-13), North America (6,14) and Singapore (15); Türkiye and other middle-income settings are absent. Turkish research has addressed general demand (2) and older-adult utilization (3), but none has profiled frequent users within the 112 system, a gap that matters because the demographic composition, primary care access, social safety net, and urban–rural balance of Turkish provinces differ from the settings that dominate the literature. Using a full calendar year of system-wide data from a mid-sized province, we investigated who the frequent users of the Karaman 112 EMS were, which factors independently predicted frequent use, and whether they formed clinically distinct subgroups requiring differentiated responses.

Materials and Methods

Study Design and Setting

This retrospective cross-sectional study examined all prehospital 112 EMS records from Karaman Province, Türkiye, for the 2025 calendar year. A single 112 command centre serves the province by dispatching ambulances across urban and rural districts. Reporting follows the STROBE guidelines for cross-sectional studies. The study was approved by the Karamanoğlu Mehmetbey University Faculty of Medicine Dean's Office, Local Scientific Medical Research Ethics Committee (approval number: 01-2026/05, date: 10.02.2026) and conducted in accordance with the Declaration of Helsinki (as revised in Brazil 2013). Because it used routinely collected, fully anonymised administrative records without patient contact, the committee waived the requirement for informed consent.

Data Source

Each record is a crew-completed patient care report for a single EMS episode from the provincial 112 information system, containing demographics, call characteristics, field assessment, field International Classification of Diseases 10th Revision, (ICD-10) diagnosis, and pre-computed operational intervals; it does not include call-taker logs. Each "call", therefore, denotes a crew-attended episode resulting in resource deployment (a response), which underlies our focus on frequent users rather than frequent callers. Patient identity was established through the Turkish national identification number, linking multiple records to the same individual.

Study Population

Of the 30,255 initial records, we excluded records with a missing national identification number (n=1,462; 4.8%; precluding linkage) or with an inter-facility transfer outcome (n=2,446; 8.1%). The final cohort comprised 26,347 records from 18,653 unique patients (Figure 1).

Definition of Frequent Users

Frequent users were defined a priori as those with five or more calls within the study year (n=430, 2.3%; 3,260 calls, 12.4%). This threshold was numerically comparable to prior prehospital studies (7,11) but more conservative, because those studies counted emergency calls to the dispatch centre whereas we counted crew-attended responses. It was fixed in advance rather than derived from the present data. The annual count aggregated the records by national identification number. Three stricter thresholds served as sensitivity definitions: ≥6 (n=282), ≥8 (n=142) and ≥10 (n=79) calls/year.

Variables

Variables included: age (continuous and dichotomised at 65 years for modelling), sex, incident area (urban vs. rural, reflecting incident location, not a verified home address), incident location (home, street/public, health care facility, prison, other), district, triage priority (red, yellow, green, black), field ICD-10 diagnosis, insurance status (retiree vs. other), and on-scene and total crew engagement intervals (seconds). Three derived indicators were defined: a primary psychiatric or substance-use diagnosis (ICD-10-chapter F) and a primary respiratory diagnosis (chapter J), both from the first-listed diagnosis, and 24-hour recontact, a call ≤24 h after the patient's preceding call; full definitions appear in the Supplementary Methods. For patient-level summaries, each patient was assigned the modal incident area and median intervals across their calls.

Statistical Analysis

Because the study analysed every eligible EMS record in the province during the study year, it constituted a complete population census, and no a priori sample-size estimation was performed. Analyses were conducted using IBM SPSS Statistics v28.0 (IBM Corp., Armonk, NY, USA); significance was set at $p<0.05$, two-tailed. Continuous variables are reported as medians with 25th-75th percentiles (interquartile range, IQR), and categorical variables as counts (%); at the call level (n=26,347), groups were compared with Mann-Whitney U and chi-square tests (Fisher exact test for expected counts <5). Missing data were minimal [326 calls (1.2%) carried no red, yellow, or green triage code: black, social-indication, or unrecorded; primary chapter unassignable in 13 calls (<0.1%)] and were retained as

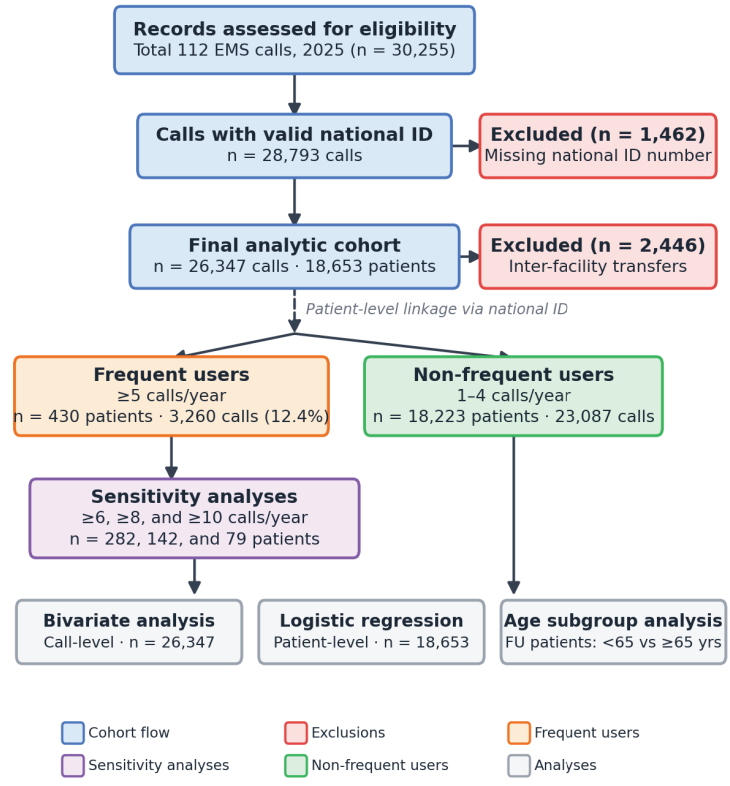


Figure 1. Study cohort flow. The initial 30,255 call records underwent sequential exclusions: records missing a national identification number (n=1,462; 4.8%) and inter-facility transfers (n=2,446; 8.1%), leaving 26,347 calls from 18,653 patients. Frequent users (≥5 calls/year): 430 patients (2.3%), 3,260 calls (12.4%). Maximum annual calls by one patient: 41

EMS: Emergency medical services, ID: Identification, FU: Frequent users

zeros. To identify independent predictors, we fitted a patient-level multivariable logistic regression model (n=18,653). Each patient was considered positive for a predictor if it was present in at least one call, so patient-level prevalences differ from call-level proportions (Table 1). Nine a priori predictors were entered simultaneously: yellow or red triage code, 24-hour recontact, psychiatric or substance-use diagnosis, respiratory diagnosis, home location, retiree insurance, urban incident area, male sex and age ≥65 years. Results are presented as adjusted odds ratios (AORs) with 95% confidence intervals (CIs); multicollinearity variance inflation factor (VIF) and model fit (Nagelkerke pseudo-R^2, Hosmer-Lemeshow) were assessed. Pre-specified secondary analyses comprised the two age-interaction terms; a full ICD-10-chapter comparison (Bonferroni-corrected, 21 chapters); separate injury (S-T) and external-cause (V-Y) flags; re-estimation under stricter thresholds; and two volume-dependence models for the triage and recontact flags: one omitting them, one with patient-clustered errors (Supplementary Methods; Tables S1, S2). The frequent-user group was further described at the patient level, stratified at 65 years of age.

Results

Cohort and Prevalence

After exclusions, 26,347 calls from 18,653 patients remained (Figure 1). Under the ≥5 calls/year threshold, 430 patients (2.3%) were frequent users, generating 3,260 calls (12.4%); the single most frequent user accounted for 41 calls.

Call-level Characteristics

At the call level, frequent-user calls differed across most domains (Table 1): patients were older (median age 64 vs 45 years), were more often male and from urban areas, and called more often from home (66.4% vs 48.2%), while respiratory diagnoses were roughly twice as common and 24-hour recontacts were over three times as common. One informative contrast is that a primary psychiatric or substance-use diagnosis was almost equally frequent in both groups (8.1% vs 7.6%; p=0.40), yet ranked among the strongest patient-level predictors.

Table 1. Characteristics of EMS calls by frequent-user status (call-level)				
Variable	**Total (n=26,347)**	**FU (n=3,260)**	**NFU (n=23,087)**	**p**
Age, median (IQR), years	47 (23-70)	64 (37-77)	45 (22-69)	<0.001
Age ≥65 years, n (%)	8,498 (32.3)	1,623 (49.8)	6,875 (29.8)	<0.001
Male sex, n (%)	13,438 (51.0)	1,922 (59.0)	11,516 (49.9)	<0.001
Urban incident area, n (%)	22,198 (84.3)	2,947 (90.4)	19,251 (83.4)	<0.001
Incident location				<0.001
Home, n (%)	13,286 (50.4)	2,164 (66.4)	11,122 (48.2)	<0.001
Street/public, n (%)	6,331 (24.0)	539 (16.5)	5,792 (25.1)	<0.001
Prison, n (%)	804 (3.1)	165 (5.1)	639 (2.8)	<0.001
Triage priority				<0.001
Red code, n (%)	531 (2.0)	47 (1.4)	484 (2.1)	0.013
Yellow code, n (%)	15,671 (59.5)	2,021 (62.0)	13,650 (59.1)	0.002
Green code, n (%)	9,819 (37.3)	1,162 (35.6)	8,657 (37.5)	0.041
Psychiatric/substance-use dx (primary F), n (%)	2,028 (7.7)	263 (8.1)	1,765 (7.6)	0.397
Respiratory dx (primary J), n (%)	1,302 (4.9)	309 (9.5)	993 (4.3)	<0.001
Retiree insurance, n (%)	1,885 (7.2)	341 (10.5)	1,544 (6.7)	<0.001
Recontact within 24 h, n (%)	927 (3.5)	315 (9.7)	612 (2.7)	<0.001
On-scene interval, median (IQR), s	563 (375-797)	618 (410-866)	555 (372–785)	<0.001
Crew engagement interval, median (IQR), s	2,181 (1,665-3,005)	2,256 (1,748-3,021)	2,170 (1,655-3,001)	<0.001

FU: Frequent users, NFU: Non-frequent users, EMS: Emergency medical services, IQR: Interquartile range, Mann-Whitney U for continuous, Pearson chi-square/Fisher exact for categorical. Triage p from 3-category test. Triage rows do not sum to 100%: the remaining 326 calls (1.2%) were assigned a black code (n=293), a social-indication code (n=18), or had no recorded code (n=15); percentages are calculated using all calls in each column as the denominator. Incident locations other than home, street/public, and prison (i.e., health care facility and other locations; n=5,926; 22.5%) are not shown. Psychiatric and respiratory conditions were defined based on the primary (first-listed) diagnosis. Call-level comparisons are descriptive and do not account for within-patient clustering; the p values are unadjusted and exploratory, and formal inference rests on the patient-level model (Table 2) and the patient-clustered call-level model (Table S2).

Independent Predictors

In the patient-level model (n=18,653), all nine predictors were independently associated with frequent use (Table 2). The two largest were a yellow or red triage code (AOR: 12.66, 95% CI: 7.30-21.96) and 24-hour recontact (12.15, 9.46-15.60), followed by psychiatric or substance-use (7.99) and respiratory (6.53) diagnoses, then home location (4.00); retiree insurance, urban area, male sex and age ≥65 years were weaker but significant. The model showed moderate explanatory power (Nagelkerke pseudo-R^2=0.372), adequate calibration (Hosmer-Lemeshow p=0.56) and no multicollinearity (all VIFs ≤1.33).

The psychiatric finding shows why the unit of analysis matters: it had no call-level association (8.1% vs 7.6%); yet, at the patient level, a primary psychiatric diagnosis was recorded in 26.7% of frequent users (115/430) versus 8.9% of others (1,614/18,223). This corresponded to an unadjusted odds ratio (OR) of 3.76 (95% CI: 3.02-4.68), rising to an AOR of 7.99 after adjustment. The shift from call-level to patient-level, rather than confounder adjustment, accounts for most of the apparent magnitude.

Diagnostic Spectrum

Across all 21 primary ICD-10 chapters (Table S3), frequent-user calls were over-represented in respiratory disease (9.5% vs 4.3%), care coordination (Z) factors, and nervous system, endocrine, and blood disorders. They were markedly under-represented among external causes (4.3% vs 17.3%) and injuries (1.9% vs 7.1%). The principal contrasts survived Bonferroni correction. Injury or external-cause presentations occurred in 6.2% of frequent-user calls versus 24.4% of other calls (Table S3) and were concentrated in younger patients (Table 3).

Sensitivity Analyses

Under stricter thresholds (≥6, ≥8, ≥10 calls/year; Table 4), the core predictors (psychiatric and respiratory diagnoses, 24-hour recontact, home location and retiree insurance) held their direction and magnitude; male sex and age ≥65 years lost significance in smaller strata, with overlapping rather than reversed CIs. Removing the two volume-entangled terms left the seven remaining predictors essentially unchanged (Table S1). In a call-level model with patient-clustered errors (Table S2), the yellow-or-red triage estimate was null (OR: 0.98,

Table 2. Adjusted predictors of frequent EMS use (n=18,653; 430 frequent users)

Predictor	AOR	95% CI	p
Yellow or red triage code	12.66	7.30-21.96	<0.001
Recontact within 24 hours	12.15	9.46-15.60	<0.001
Psychiatric/substance-use dx (F)	7.99	6.01-10.63	<0.001
Respiratory dx (J)	6.53	4.96-8.59	<0.001
Home incident location	4.00	2.91-5.49	<0.001
Retiree insurance	3.28	2.51-4.27	<0.001
Urban incident area	2.27	1.55-3.33	<0.001
Male sex	1.51	1.20-1.90	<0.001
Age ≥65 years	1.40	1.08-1.83	0.011

AOR: Adjusted odds ratio, CI: Confidence interval, VIF: Variance inflation factor. EMS: Emergency medical services, patient-level multivariable, logistic regression. Nagelkerke pseudo-R^2 =0.372. All VIFs ≤1.33. Hosmer-Lemeshow goodness-of-fit test: χ^2=6.77, df=8, p=0.56 (no evidence of poor fit). Effect modification (separate model with interaction terms): age ≥65 × psychiatric AOR: 0.81 (95% CI: 0.41-1.61, p=0.54); age ≥65 × respiratory AOR: 1.14 (95% CI: 0.65-1.99, p=0.64). Neither significant

p=0.71), confirming that its large patient-level association arises from accumulation across repeated calls rather than from per-encounter acuity. By contrast, 24-hour recontact retained the strongest call-level association (OR: 4.59), and respiratory diagnosis and home location also remained independently associated (OR: 2.10 and 1.77), indicating genuine per-encounter characteristics rather than volume artefacts.

Age-stratified Profiles

When analyzed at the patient level and stratified at 65 years (Table 3), the 430 frequent users formed two clinically distinct groups. Younger users (<65 years, n=213) more often had a psychiatric or substance-use diagnosis (44.6% vs 9.2%), street or public encounters (56.8%) and prison encounters (12.2%), and 24-hour recontact. Older users (≥65 years, n=217) called almost exclusively from home (95.9%), more often had a respiratory diagnosis (36.9%), nearly all received a yellow or red triage code (99.5%), more often held retiree insurance (63.1%) and needed longer intervals. Neither the age × psychiatric (p=0.54) nor the age × respiratory (p=0.64) interaction was significant; therefore, these differences reflect variation in prevalence rather than age-dependent effect sizes.

Temporal Distribution

The monthly distribution did not differ between groups (Figure S1; p = 0.37), nor did shift or day-of-week patterns.

Discussion

In this province-wide, year-long dataset of 112 EMS records, 2.3% of patients accounted for 12.4% of calls. The concentration is consistent with structurally different systems: Kjærgaard et al. (11) found the top 1% of Danish callers corresponded to roughly five or more calls per year (the same ≥5 threshold identified 2.3% of patients here), and Viken et al. (7) reported 3.2% of Bergen callers generated about one in four calls.

Table 3. Age-stratified profiles of the 430 frequent users (patient-level)

Variable	<65 years (n=213)	≥65 years (n=217)	p
Male sex, n (%)	122 (57.3)	137 (63.1)	0.215
Urban incident area (modal), n (%)	203 (95.3)	191 (88.0)	0.006
Location: home, n (%)	168 (78.9)	208 (95.9)	<0.001
Location: street/public, n (%)	121 (56.8)	62 (28.6)	<0.001
Location: prison, n (%)	26 (12.2)	3 (1.4)	<0.001
Psychiatric/substance-use dx, n (%)	95 (44.6)	20 (9.2)	<0.001
Respiratory dx (primary J), n (%)	42 (19.7)	80 (36.9)	<0.001
Yellow or red triage, n (%)	200 (93.9)	216 (99.5)	<0.001
Retiree insurance, n (%)	28 (13.1)	137 (63.1)	<0.001
Recontact within 24 h, n (%)	95 (44.6)	66 (30.4)	0.002
Injury dx (S-T), n (%)	33 (15.5)	13 (6.0)	0.001
External-cause dx (V-Y), n (%)	62 (29.1)	34 (15.7)	<0.001
Any injury/external-cause, n (%)	77 (36.2)	43 (19.8)	<0.001
On-scene interval, median (IQR), s	520 (408-720)	712 (588-861)	<0.001
Crew engagement interval, median (IQR), s	2,031 (1,684-2,489)	2,430 (2,095-2,872)	<0.001

IQR: Interquartile range, patient-level; each patient counted once. Binary flags = recorded in ≥1 call; location and diagnosis categories within a domain are not mutually exclusive (a patient with calls from both home and the street is counted in both rows), so within-domain percentages may exceed 100%. Mann-Whitney U/chi-square/Fisher exact as appropriate; all comparisons are descriptive, with unadjusted p values interpreted as exploratory

Table 4. Sensitivity analysis across frequency thresholds adjusted odds ratio				
Predictor	**≥5 (main, n=430)**	**≥6 (n=282)**	**≥8 (n=142)**	**≥10 (n=79)**
Yellow or red triage code	12.66	11.04	14.49	20.82
Recontact within 24 hours	12.15	11.79	15.28	13.81
Psychiatric/substance-use dx (F)	7.99	8.81	8.08	7.50
Respiratory dx (J)	6.53	7.62	7.62	9.62
Home incident location	4.00	3.53	4.64	4.71
Retiree insurance	3.28	3.15	3.03	4.91
Urban incident area	2.27	2.87	3.02	2.47
Male sex	1.51	1.48	1.34	0.89
Age ≥65 years	1.40	1.34	1.53	1.35
Nagelkerke: 0.372 (≥5), 0.37 (≥6), 0.398 (≥8),0.412 (≥10). At ≥10, the triage estimate is unstable (95% confidence interval: 2.83-153.3) owing to near-complete separation. Male sex and age ≥65 lose significance at stricter thresholds				

The two largest patient-level associations are not equivalent: yellow or red triage codes reflect repeated contact rather than per-encounter severity (null at the call level, Table S2), whereas 24-hour recontact retains a genuine per-encounter signal even as it rises with call volume. The primary psychiatric and respiratory diagnoses are the underlying clinical drivers.

Nearly one in ten frequent-user calls occurred within 24 hours of a preceding call (9.7% vs 2.7%), the second-largest adjusted association, a pattern Kjærgaard et al. (11) also reported. Structural features of the provincial Turkish healthcare system plausibly channel these early returns toward ambulance services. Primary care rests on family-medicine practices that operate largely during weekday daytime hours, and hospital care requires no referral. After hours, the free 112 line is the only telephone access to medical help; there is no counterpart to the out-of-hours services that absorb low-acuity demand in the Norwegian and Danish systems (10,11). At the receiving emergency departments, heavy caseloads can shorten observation and limit structured discharge guidance. When symptoms persist or anxiety recurs overnight, a further 112 call may follow. Each subgroup faces a different service gap: older users with respiratory disease are not routinely offered early follow-up after discharge, whereas younger users in psychiatric crisis lack round-the-clock community mental health services, leaving the emergency department as the only after-hours option. Because the records capture neither the reason for a recontact nor the disposition of the preceding encounter, these mechanisms remain hypotheses.

Among our frequent users, psychiatric diagnoses and male sex were both prominent, as described in Norway (7), Denmark (1) and the United States (6); in the US cohort, frequent use concentrated among younger and middle-aged patients (6). Respiratory diagnoses were more common in older frequent users (36.9% vs 19.7%) and independently associated with frequent use (AOR: 6.53), echoing Søvsø et al. (1). The longer intervals of older users fit recurrent chronic obstructive pulmonary disease exacerbations, the commonest final diagnosis among repeated EMS users (16), for whom proactive follow-up has been proposed (14). Across all chapters, repeated use reflected chronic and care-coordination needs rather than injury; residual trauma was concentrated mainly among younger users.

At the patient level, the group split at 65 years: younger users were characterized by psychiatric diagnoses and social vulnerability (encounters on the street and in prison), whereas older users were characterized by chronic respiratory disease. A younger-social, older-medical division was also reported in Singapore (15). Interaction testing showed that the strength of each association did not differ by age; therefore, the profiles reflect differences in prevalence across age groups rather than age-dependent effects.

Frequent use concentrated in home settings within urban areas (home 66.4% vs 48.2%; urban 90.4% vs 83.4%; urban an independent predictor, AOR: 2.27), consistent with Pakdemirli et al. (3), who found 88.6% of elderly İzmir calls urban. Retiree insurance independently predicted frequent use (AOR: 3.28) even after age adjustment. Because 112 transport is free, this likely reflects a pensioner population with comprehensive coverage and a high chronic disease burden rather than insurance-specific access. Its concentration among older home-based users (63.1%) is consistent with an access-facilitated pattern in which a free, appointment-free service lowers the threshold for using the ambulance as a route to care. Whether social isolation also contributes, given that 38.8% of Canadian frequent callers live alone (17), lies beyond what can be determined from administrative data.

Unlike general Turkish EMS demand, where Sariyer et al. (2) documented daytime–evening variation, our groups showed no temporal difference (Figure S1).

Strengths

To our knowledge, this is the first patient-level characterisation of frequent users within the Turkish 112 system, a middle-income setting not covered by the existing literature. Linkage of repeated contacts to the same individual through the national identification number lets us separate per-encounter associations from cumulative associations. The a priori threshold also held under three stricter cut-offs and in a patient-clustered call-level model (Table S2).

Clinical Implications

These subgroups point to distinct service responses: psychiatric liaison and social support for younger users, and coordinated chronic-disease and geriatric care for older users, the kind of case management that has reduced frequent use elsewhere (13). Within the Turkish 112 framework, such responses require system-level change, as ambulances currently convey patients almost exclusively to emergency departments. For services developing automated frequent-user flags, a note of caution: the yellow-or-red triage code, although the largest patient-level association, carried no per-encounter signal (Table S2) and would mislabel these patients if applied predictively. Cumulative contact count and clinical phenotype are more reliable.

Research Implications

Multi-centre replication across Turkish provinces is needed before these findings can inform policy. Whether prospective identification of frequent users, coupled with coordinated follow-up, can reduce avoidable activations is testable but unproven and represents the natural next step. Linking EMS records to the social, housing, and clinical data that administrative systems omit would clarify the upstream drivers that this study can only outline. Linkage with hospital electronic health records, adding emergency department disposition, admission, and short-term outcomes, would allow direct evaluation of the clinical appropriateness and, ultimately, the preventability of these transports.

Study Limitations

This was a retrospective, single-centre, single-year study. Prevalence was calculated using EMS users, rather than provincial residents, as the denominator; therefore, per-capita rates were not computed. The 4.8% of records excluded for missing identification numbers may over-represent socially isolated, homeless, or undocumented individuals who are linked to elevated EMS use and behavioural health needs (6,17), thereby underestimating both the prevalence of frequent use and the psychiatric association. ICD-10 codes reflect field impressions rather than confirmed diagnoses, and socioeconomic status, housing, social isolation and comorbidity measures, associated with frequent use elsewhere (17,18), were unavailable. The records do not identify who placed each call; including third-party calls roughly doubled frequent contacts in one Norwegian system (10). Because a patient was counted as positive whenever an indicator appeared in any call, patients with higher call volumes had more opportunities to screen positive for every binary predictor; this effect was most pronounced for the triage code, which showed a large patient-level association that was absent at the call level (Tables 1 and S2). Although the composition of frequent users may vary across provinces, the 112 system operates under the same national organisation, command-centre structure, and free-transport policy in every province of Türkiye, which supports cautious transferability of the system-level findings; the multi-centre replication called for above remains the definitive test. The cross-sectional design precludes causal inference, and post-encounter outcomes were unavailable, limiting assessment of clinical appropriateness.

Conclusion

Frequent users of the Karaman 112 EMS system are a small yet disproportionately demanding group: 2.3% of patients generated 12.4% of calls; they constitute two clinically distinct subpopulations rather than one. The largest patient-level signal, a yellow or red triage code, reflects accumulated contact rather than per-encounter acuity, while primary psychiatric and respiratory diagnoses identify the clinical drivers.

Ethics

Ethics Committee Approval: The study was approved by the Karamanoğlu Mehmetbey University Faculty of Medicine Dean's Office, Local Scientific Medical Research Ethics Committee (approval number 01-2026/05, dated 10.02.2026) and conducted in accordance with the Declaration of Helsinki (as revised in Brazil 2013).

Informed Consent: This is retrospective cross-sectional study.

Footnotes

Authorship Contributions

Concept: S.Ö., Design: S.Ö., Data Collection or Processing: S.Ö., M.S.Y., Analysis or Interpretation: S.Ö., M.S.Y., Literature Search: S.Ö., Writing: S.Ö., M.S.Y.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

Supplementary Link: https://d2v96fxpocvxx.cloudfront.net/d363ec1e-9e5e-4591-a00a-d656bfcabb80/content-images/91b0de03-b3af-47bf-b99c-dc45706e8876.pdf

References

1. Søvsø MB, Kløjgaard TA, Hansen PA, Christensen EF. Repeated ambulance use is associated with chronic diseases – a population-based historic cohort study of patients' symptoms and diagnoses. Scand J Trauma Resusc Emerg Med. 2019;27:46.
2. Sariyer G, Ataman MG, Akay S, Sofuoglu T, Sofuoglu Z. An analysis of emergency medical services demand: time of day, day of the week, and location in the city. Turk J Emerg Med. 2017;17:42-7.
3. Pakdemirli A, Bayram B, Güvenç E, Ellidokuz H. Evaluation of ambulance calls for patients over 65 years of age in İzmir, Turkey: a two-year retrospective analysis. Turk J Med Sci. 2021;51:1253-60.
4. Özata M, Toygar Ş, Yorulmaz M, Cihangiroğlu N. Comparative analysis of using 112 emergency ambulance services in Turkey and the province of Konya. Eur J Gen Med. 2011;8:262-7.
5. Scott J, Strickland AP, Warner K, Dawson P. Frequent callers to and users of emergency medical systems: a systematic review. Emerg Med J. 2014;31:684-91.
6. Knowlton A, Weir BW, Hughes BS, Southerland RJ, Schultz CW, Sarpatwari R, et al. Patient demographic and health factors associated with frequent use of emergency medical services in a midsized city. Acad Emerg Med. 2013;20:1101-11.
7. Viken SN, Myrmel L, Brattebø G. Frequent callers contacting the Norwegian national emergency medical number 113: a retrospective study. Scand J Trauma Resusc Emerg Med. 2024;32:102.
8. Alm F, Kurland L, Hugelius K. Frequent users of the ambulance service in a Swedish region: a retrospective cohort study. Scand J Trauma Resusc Emerg Med. 2026;34:16.
9. Harring AKV, Kjærgaard M, Gehrt TB. Frequent callers vs. frequent users – a scoping review of frequent contacts to the emergency medical services. Int J Emerg Med. 2025;18:108.
10. Harring AKV, Graesli O, Häikiö K, Hjortdahl M, Jørgensen TM. Frequent contacts to emergency medical services (EMS): more than frequent callers. BMC Emerg Med. 2024;24:190.
11. Kjærgaard M, Harring AKV, Gehrt TB. How to define a frequent caller to the prehospital emergency medical services? Literature-based vs. data-driven approach. Prehosp Emerg Care. 2026;30:474-83.
12. Scott J, Strickland AP, Warner K, Dawson P. Describing and predicting frequent callers to an ambulance service: analysis of 1 year call data. Emerg Med J. 2014;31:408-14.
13. Edwards MJ, Bassett G, Sinden L, Fothergill RT. Frequent callers to the ambulance service: patient profiling and impact of case management on patient utilisation of the ambulance service. Emerg Med J. 2015;32:392-6.
14. Norman C, Mello M, Choi B. Identifying frequent users of an urban emergency medical service using descriptive statistics and regression analyses. West J Emerg Med. 2016;17:39-45.
15. Kuek BJW, Li H, Yap S, Ng MXR, Ng YY, White AE, et al. Characteristics of frequent users of emergency medical services in Singapore. Prehosp Emerg Care. 2019;23:215-24.
16. Tärnqvist J, Dahlén E, Norberg G, Magnusson C, Herlitz J, Strömsöe A, et al. On-scene and final assessments and their interrelationship among patients who use the EMS on multiple occasions. Prehosp Disaster Med. 2017;32:528-35.
17. Agarwal G, Lee J, McLeod B, Mahmuda S, Howard M, Cockrell K, et al. Social factors in frequent callers: a description of isolation, poverty and quality of life in those calling emergency medical services frequently. BMC Public Health. 2019;19:684.
18. Pouw MA, Dijkstra H, Kőntős ER, de Jonghe A, Dercksen B, Maruster L, et al. Frequent emergency medical services utilization among older patients: exploration and automatic identification using natural language processing. J Am Med Dir Assoc. 2025;26:105814.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.23356
Eurasian J Emerg Med. 2026;25: 411-8

Evaluating the Performance of Large Language Models in Radiological Decision-making: A Comparative Study Using ACR Appropriateness Criteria for Acute Abdomen

Abdullah Enes Ataş[1], Halil İbrahim Şara[1], Fatih Cemal Tekin[2]

[1]Necmettin Erbakan University Faculty of Medicine, Department of Radiology, Konya, Türkiye
[2]Konya City Hospital, Clinic of Emergency Medicine, Konya, Türkiye

Abstract

Aim: To evaluate the accuracy of four large language models in assessing the appropriateness of imaging tests for patients with suspected acute abdomen, using the American College of Radiology (ACR) Appropriateness Criteria as the gold standard.

Materials and Methods: This cross-sectional study evaluated ChatGPT 5.5, Gemini 3.1 Pro, Claude Opus 4.8, and DeepSeek V4 Pro across 30 clinical scenarios of acute abdominal pain. Models were tested using zero-shot and one-shot prompting strategies, yielding a total of 240 responses. Responses were compared to the ACR categories of "Usually Appropriate, "May Be Appropriate," and "Usually Not Appropriate." Performance was measured using exact match accuracy, linearly weighted Cohen's kappa, and major discordance rates.

Results: The overall accuracy of the pooled models was 86.3%, with a weighted kappa of 0.78. Claude Opus 4.8 achieved the highest accuracy at 95.0% and a kappa of 0.91. Clinically risky major discordance rates remained low across all models, ranging from 3.3% to 8.3%. Utilizing a one-shot prompting strategy increased overall accuracy from 81.7% to 90.8% compared to zero-shot prompting.

Conclusion: Large language models successfully align imaging recommendations for acute abdominal conditions with established ACR guidelines. Providing a guideline-formatted example prior to the task improves model performance and highlights their potential as reliable radiological decision-support tools. By supporting faster, guideline-concordant imaging selection, such tools may help reduce unnecessary ionizing radiation exposure and alleviate decision-making pressure in high-volume emergency settings.

Keywords: Radiology, acute abdomen, artificial ıntelligence, clinical decision support systems, natural language processing

Introduction

Acute abdominal pain is a very broad and complex clinical condition that physicians most frequently encounter in emergency departments and outpatient clinics of hospitals, with underlying causes ranging from simple and self-limiting disorders to life-threatening conditions requiring emergency surgical intervention (1). While the patient's history, physical examination, and basic laboratory tests help physicians narrow down the list of possible diseases, radiological imaging methods are often needed to make a definitive diagnosis and create the correct treatment plan (2). However, choosing the right imaging method within this wide range of differential diagnoses is not always easy. Ordering unnecessary or clinically inconsistent imaging tests exposes patients to ionizing radiation, increases healthcare system costs, and slows diagnostic processes by causing inefficient use of limited resources in emergency departments (3,4).

To overcome these challenges and ensure standardization in clinical decision-making, the American College of Radiology (ACR) developed the "Appropriateness Criteria," an evidence-based guideline (1,5). These criteria, updated annually by expert panels using current data from the medical literature, guide physicians in selecting the most appropriate radiological imaging or treatment method for a given clinical scenario (5,6). In fact, in the United States, under regulations aimed at improving the

Corresponding Author: Abdullah Enes Ataş MD, Necmettin Erbakan University Faculty of Medicine, Department of Radiology, Konya, Türkiye
E-mail: aenesatas@gmail.com **ORCID ID:** orcid.org/0000-0001-6623-3024

Cite this article as: Ataş AE, Şara Hİ, Tekin FC. Evaluating the performance of large language models in radiological decision-making: a comparative study using ACR appropriateness criteria for acute abdomen. Eurasian J Emerg Med. 2026;25: 411-8

Received: 14.07.2026
Accepted: 16.08.2026
Published: 04.09.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

quality of healthcare, consulting such clinical decision support systems is encouraged when ordering advanced imaging studies (3,6). Despite this, the use of these guidelines in daily and intensive clinical practice remains generally limited. The main reason for this is that establishing a direct link between complex patient notes kept in free text and guidelines with strict rules and structured templates is time-consuming and difficult (7).

This gap between guideline complexity and everyday practice is most pronounced in the emergency department, where physicians must select an imaging strategy within minutes, under diagnostic uncertainty, and often outside regular working hours when subspecialty radiology consultation may be less readily available. In this setting, an unnecessary computed tomography scan not only exposes the patient to avoidable ionizing radiation but also prolongs the patient's length of stay in the emergency department and increases downstream imaging costs. A decision-support tool capable of instantly cross-referencing a free-text clinical presentation against ACR-based recommendations could, therefore, help emergency physicians select the correct initial imaging test more consistently, reduce unnecessary radiation exposure, and support timely, evidence-based triage decisions.

Recent advancements in artificial intelligence, particularly in large language models (LLMs), hold significant potential to address this problem (8). These systems, trained to understand and generate human language, promise the ability to rapidly analyze patient data in free-text format and function as evidence-based clinical decision support systems (8). Research in the medical literature indicates that LLMs show promise in radiology for interpreting clinical scenarios and recommending examinations. Furthermore, it has been found that an AI model specifically supported by ACR guidelines can outperform radiologists and standard language models in selecting examinations (9). Similarly, it has been proven that LLMs can predict the correct imaging tests with high agreement in various radiology subspecialties (10). However, given the sensitivity of medical data, the risk of hallucinations and the capacity of these models to meet clinical safety standards remain obstacles to their widespread adoption (11). While the current medical literature has demonstrated the potential of LLMs for general radiological decision-making, direct comparative studies assessing the extent to which the newest, most advanced language models align with the ACR Appropriateness Criteria in field-specific scenarios are lacking.

The aim of this research is to address this gap in the literature by evaluating how accurately four current LLMs, which are increasingly used in the medical field, can assess the appropriateness of imaging tests for patients with suspected acute abdomen.

Materials and Methods

Study Design

This cross-sectional, comparative study is designed to measure the ability of LLMs to assess the appropriateness of radiological examinations. The main focus of the study is to determine the extent to which the responses generated by four LLMs regarding the appropriateness of a selected imaging test for a patient presenting with acute abdominal pain are consistent with reference guidelines. All clinical scenarios were completed using the most current versions of the LLMs dated between June 1 and July 1, 2026.

The ACR Appropriateness Criteria® published by the ACR were used as the reference and gold standard in the study. These criteria, prepared by expert panels, evaluate each imaging or treatment option for a given clinical scenario on a comprehensive 9-point scale. For ease of use in daily practice, this scale is generally divided into three categories: "Usually Appropriate" (green category), "May Be Appropriate" (yellow category), and "Usually Not Appropriate" (red category). Throughout the study, the accuracy of the models' responses was tested by comparing them to the gold standard template consisting of these three categories.

Ethical Considerations

This study was completed using current versions of publicly available LLMs, drawing on ACR's publicly available guidelines (12). Because no patient data, hospital records, or images were used, approval from the ethics committee was not required.

Clinical Scenarios

For the study, thirty (30) clinical case scenarios were prepared, directly corresponding to specific variants defined under the heading of acute abdominal pain in the ACR Appropriateness Criteria. These scenarios were selected to cover nine basic clinical presentations frequently encountered by physicians in daily clinical and emergency practice. These are: non-localized generalized abdominal pain, right lower quadrant pain (suspected appendicitis), right upper quadrant pain, left lower quadrant pain (suspected diverticulitis), left upper quadrant pain, epigastric pain, suspected small bowel obstruction, suspected acute mesenteric ischemia, and acute pyelonephritis.

For each prepared scenario, a specific imaging test (e.g., non-contrast computed tomography or abdominal ultrasonography) was selected, and the correct ACR guideline category for this scenario-test pairing was recorded as the gold-standard response. The 30 selected scenarios were balanced to reflect that, in actual clinical guidelines, the tests deemed appropriate outnumber

those deemed inappropriate. In this context, 17 of the scenarios were created from test matches belonging to the "Usually Appropriate" category, 6 from "May Be Appropriate", and 7 from "Usually Not Appropriate."

All 30 vignettes were drafted by a board-certified radiologist with 10 years of experience and were based directly on the illustrative clinical scenarios provided in the corresponding ACR Appropriateness Criteria variant. Each vignette and its assigned gold-standard category were then independently checked against the published ACR document by a second radiologist with 8 years' experience and an emergency medicine physician; disagreements were resolved by consensus. The vignettes were written as concise, single-issue presentations rather than as verbatim excerpts from real patient records, since no patient data were used in this study; this simplification is discussed as a limitation.

LLMs and Prompting Strategy

All models were accessed via a web interface between June 1 and July 1, 2026, using the following snapshots: ChatGPT 5.5, Gemini 3.1 Pro, Claude Opus 4.8, and DeepSeek V4 Pro. Default vendor settings were used for all models. Each of the 60 prompts per model (30 scenarios × 2 prompting strategies) was submitted as a new, independent conversation with no shared context, to ensure that no model's response to one scenario was influenced by its response to a previous one.

Four advanced LLMs were selected for evaluation in this study: ChatGPT 5.5, Gemini 3.1 Pro, Claude Opus 4.8, and DeepSeek V4 Pro. Each model was presented with 30 prepared case scenarios, using two different prompting strategies. This allowed each model to respond to a total of 60 queries (30 scenarios × 2 prompting strategies). The two strategies used to guide the models are:

• Zero-shot prompting: In this method, the model is given only a the task brief instruction explaining, followed by the clinical scenario and the suggested imaging test that it must evaluate directly. No examples of what the expected response should be are shown to the model beforehand.

• One-shot prompting: In this method, the model is given a single task instruction and a solved example is shown before the actual questions. This example illustrates that the study included a scenario outside the 30 cases, a test, the category deemed correct according to the guideline, and a brief justification for that classification. Thus, the model was able to see exactly what format and logical framework for the response were expected of it before proceeding to the actual task.

Across all methods employed, models were asked to provide responses in one of the three specified categories ("Usually Appropriate", "May Be Appropriate", or "Usually Not Appropriate") and a brief justification consisting of one or two sentences supporting their preference. As a result, 240 model responses were obtained by presenting 30 scenarios to 4 models using 2 strategies. Each of these responses was independently compared with the gold standard ACR category to which it belonged.

The complete wording of the fixed instruction, the one-shot worked example, and the clinical scenario and imaging study text for all 30 cases are provided in Supplementary Material S1, allowing independent replication of the prompting procedure

Evaluation of Responses

Each of the 240 responses obtained from the models was scored at various levels by comparing it to the established gold standard.

• Perfect match (accuracy): Whether the category suggested by the model exactly matched the correct category in the ACR guideline (yes/no) was evaluated.

• Ordered distance (absolute ordered error): Categories were assigned numerical values according to severity (Red = 1, Yellow = 2, Green = 3). The absolute difference between the numerical value chosen by the model and the gold standard value was calculated. A value of 0 indicates a perfect match; a value of 1 indicates that the model deviated by only one step (e.g., evaluating a test that is "Usually Appropriate" as "May Appropriate"); and a value of 2 indicates that the model chose a completely opposite category.

• Major discordance is the most clinically dangerous type of error, with the potential to cause the greatest harm to the patient. This occurs when the model classifies an examination as "Usually Appropriate" when it should be "Usually Not Appropriate," or vice versa. This type of error has been analyzed separately from the overall accuracy assessment, as it can result in a patient being subjected to an unnecessary test or being denied an urgent test.

Statistical Analysis

There is a natural ranking and a significance level among the response categories (red, yellow, green) in the study. An error of one category is clinically more acceptable than an error of two categories (exact contrast). Therefore, the linearly-weighted Cohen's kappa test was used as the primary indicator of agreement between each model and the ACR gold standard. Unlike simple accuracy percentages, the weighted kappa is considered a more reliable metric because it statistically corrects for chance agreement and awards the model partial credit in situations where it "narrowly missed" (e.g., predicting yellow instead of green). The Landis and Koch criteria, considered standard in this field, were used to interpret the obtained kappa

values. Accordingly, scores were categorized as follows: below 0, weak; 0.00-0.20, weak; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; and 0.81-1.00, almost perfect. For kappa analyses, 95% confidence intervals (CIs) were calculated using a bootstrap method in which the data were resampled 2,000 times.

In addition to complex statistics, the simple accuracy rate (percentage of exact matches) and the mean absolute rank error were also presented as descriptive measures to facilitate interpretation. In inter-method comparisons, the McNemar's exact test, which is suitable for paired data since the same scenarios were evaluated twice by the same model, was used to test the significance of the difference between zero-shot and one-shot methods. To test differences in success among the four models, Cochran's Q test (a paired-group test) was used, while the chi-square test was used to compare large discordance between models.

In clinical practice, preventing unnecessary imaging and avoiding missed diagnoses are vital. Therefore, the capacity of each model to correctly identify "Usually Not Appropriate" (red) tests (sensitivity) and its ability to avoid mistakenly labeling an appropriate test as inappropriate (specificity) were also calculated using the binary classification method. All data were analyzed using python (the pandas, SciPy, and statsmodels libraries), and statistical significance was defined as a two-sided p-value <0.05.

Results

Overall Agreement with ACR Goodness-of-Fit Criteria

When the total of 240 model responses obtained in the study was evaluated as a pooled sample, the overall perfect agreement (accuracy) rate between the LLMs and the ACR gold standard was 86.3% (207 of 240 responses). The linearly weighted overall kappa was 0.78 (95% CI, 0.70-0.85), corresponding to a clinically substantial level of agreement according to the Landis and Koch criteria. The mean absolute rank error was low across all analyzed responses. These data show that even when the models fail to match the ACR category perfectly, they generally deviate only into a neighboring category (one unit away) rather than giving a completely opposite response.

Inter-model Comparison

When the overall capacity of each model was examined by combining results from different request methods, the accuracies of the four models ranged from 81.7% to 95.0% (Table 1). The Claude Opus 4.8 model exhibited the highest performance, being the only model to fall into the "almost perfect" fit band with 95.0% accuracy and a weighted kappa value of 0.91 (95% CI: 0.77-1.00). The other three models, ChatGPT 5.5, Gemini 3.1 Pro, and DeepSeek V4 Pro, showed significant clinical agreement with kappa values of 0.72, 0.70, and 0.79, respectively.

Despite these observed percentage differences, no significant difference in accuracy was found when the models were tested using the Cochran Q test. Cochran's Q test showed no statistically significant difference in accuracy among the four models for either prompting condition (zero-shot: $p=0.082$; one-shot: $p=0.958$). This statistical result indicates that the apparent percentage differences in performance between the models may be attributable to the relatively small sample size in this study and should therefore be interpreted cautiously.

Examination of the "major discordance" rates, which indicate the frequency of clinically risky decisions made by the models, showed that all models kept these rates quite low. This rate was 3.3% for the Claude Opus 4.8 and DeepSeek V4 Pro, while it reached its highest value of 8.3% for the Gemini 3.1 Pro. The chi-square test revealed no statistically significant difference in the frequency of these clinical errors between the models ($p=0.533$).

The Effect of the Prompt Strategy

One important finding of the study is that the way the question is asked affects model performance. Presenting the models with a previously solved example (single-shot method), instead of simply describing the task, significantly increased the success rate compared to presenting no example (zero-shot method).

Table 1. Overall performance metrics of LLMs: Exact match accuracy, weighted Kappa, and Levels of agreement (combined zero-shot and one-shot prompting, n=60 responses per model)

Model	Exact match (accuracy)	Weighted kappa (95% CI)	Agreement level	Major discordance
ChatGPT 5.5	83.3%	0.72 (0.54-0.87)	Substantial	6.7%
Gemini 3.1 Pro	81.7%	0.70 (0.51-0.85)	Substantial	8.3%
Claude Opus 4.8	95.0%	0.91 (0.77-1.00)	Almost perfect	3.3%
DeepSeek V4 Pro	85.0%	0.79 (0.64-0.91)	Substantial	3.3%

CI: Confidence interval, LLMs: Large language models, agreement level based on Landis and Koch criteria. Major discordance = a two-category error ("Usually Appropriate" scored as "Usually Not Appropriate," or the reverse)

When responses from all models were combined, overall accuracy increased from 81.7% in the zero-shot setting to 90.8% in the one-shot setting. Similarly, the overall kappa agreement increased from 0.71 ("significant") to 0.85 ("almost perfect").

Examination of individual models indicated that this positive effect was consistently observed in three of the four models (Table 2). The McNemar test, applied to paired data, approached but did not reach the statistical significance threshold of $p<0.05$ (observed $p=0.071$). This situation reveals that a single example confers a real structural benefit to the models, but larger sets of scenarios are needed to statistically validate this benefit. Only the Claude Opus 4.8 model performed slightly better when asked without an example than when asked with an example; however, this model remained within a very high fit range under both methods.

Notably, both Cochran's Q results indicate that one-shot prompting narrowed the performance gap between models: correct-answer counts under zero-shot prompting ranged from 22 to 29 out of 30 across the four models ($Q=6.69$, $p=0.082$), whereas under one-shot prompting, this range narrowed to 27-28 out of 30 ($Q=0.31$, $p=0.958$). This suggests that providing a single worked example not only improved average accuracy but also reduced variability between different LLMs.

Identifying Inappropriate and Appropriate Tests

In clinical practice, failing to identify a test in the "Usually Not Appropriate" category can result in unnecessary testing, while failing to recommend a "Usually Appropriate" test can lead to missed diagnoses. To examine the safety profile at these two extremes, the sensitivity and specificity values of the models were calculated (Table 3). All four models correctly identified most inappropriate tests, with sensitivity rates ranging from 79% (ChatGPT 5.5) to 100% (Claude Opus 4.8). Specificity—the ability to avoid mistakenly labeling an appropriate test as "inappropriate"—was high, with models' performance ranging from 89% to 96%.

For "Usually Appropriate" tests, the calculated sensitivity values ranged from 82% to 91%, and the specificity values ranged from 88% to 100%. Claude Opus 4.8 demonstrated the best performance across all safety thresholds in both extreme scenarios.

Table 2. Comparative analysis of model performance by prompting strategy: evaluating accuracy and weighted Kappa for zero-shot versus one-shot approaches (n=30 responses per cell)

Model	Prompting strategy	Accuracy	Weighted kappa
ChatGPT 5.5	Zero-shot	76.7%	0.61
ChatGPT 5.5	One-shot	90.0%	0.84
Gemini 3.1 Pro	Zero-shot	73.3%	0.59
Gemini 3.1 Pro	One-shot	90.0%	0.81
Claude Opus 4.8	Zero-shot	96.7%	0.92
Claude Opus 4.8	One-shot	93.3%	0.89
DeepSeek V4 Pro	Zero-shot	80.0%	0.73
DeepSeek V4 Pro	One-shot	90.0%	0.85

Claude Opus 4.8 was the only model that performed marginally better with zero-shot prompting than with one-shot prompting, although both results indicate very good agreement

Table 3. Diagnostic reliability of AI models: sensitivity and specificity in classifying "Usually Not Appropriate" and "Usually Appropriate" clinical guideline categories

Model	"Not Appropriate" sensitivity	"Not Appropriate" specificity	"Appropriate" sensitivity	"Appropriate" specificity
ChatGPT 5.5	79%	96%	82%	88%
Gemini 3.1 Pro	86%	89%	82%	88%
Claude Opus 4.8	100%	96%	91%	100%
DeepSeek V4 Pro	93%	93%	82%	96%

ACR: American College of Radiology, "Not Appropriate" = ACR "Usually Not Appropriate" category; "Appropriate" = ACR "Usually Appropriate" category. Values are one-versus-rest sensitivity/specificity for each model detecting that category correctly

Distribution According to Clinical Presentations

The fit of the models to the ACR standards was not equally distributed across the nine clinical scenarios considered. Descriptive data, which were not subjected to formal statistical testing because of the small sample size, indicate that the models achieved their greatest success in cases with very clear diagnostic boundaries, such as suspected acute mesenteric ischemia (100%) and left upper quadrant pain (93.8%). In contrast, a significant decrease in model accuracy was observed in scenarios such as suspected small bowel obstruction (79.2%) and right lower quadrant pain/suspected appendicitis (80.0%), which present numerous gray areas in clinical practice. This table appears directly related to the clarity of disease markers in the models' training data.

Major-discordance responses were unevenly distributed across scenarios: 13 of the 240 responses (5.4%) met this definition and clustered in three clinical panels—right lower quadrant pain/suspected appendicitis, right upper quadrant pain, and epigastric pain/pancreatitis—which together accounted for 8 of the 13 major-discordance responses. These are precisely the presentations in which the correct ACR category depends on secondary features (pregnancy status, presence of a complication, intensive care unit setting, an equivocal prior ultrasound) rather than on the primary symptom alone, suggesting that current models are more likely to overlook a qualifying clinical detail than to misjudge the primary presentation itself. The complete case-by-case response matrix for all 30 scenarios and all four models is provided in Supplementary Material S2.

Discussion

The findings of this research reveal that AI-based LLMs have reached a highly competent level in interpreting complex medical texts and matching them with evidence-based clinical guidelines. The overall accuracy rate of 86.3% and the weighted kappa score of 0.78, both achieved by pooling data from four different language models, demonstrate that the models successfully placed a patient history in text format within the strict clinical boundaries set by the ACR.

As noted in extensive studies by Zaki et al. (10) or reports by Rau et al. (9), supporting physicians with an algorithmic system when ordering radiological examinations both increases diagnostic success and prevents unnecessary examination costs. The available data also reinforce the claims in the medical literature that these current language models provide a solid infrastructure for their integrated use in radiology clinical decision support systems. Detailed comparison of the data across models shows that all systems exceed a certain level of clinical competence. Statistical tests did not show dramatic differences between the models. Nevertheless, it is noteworthy that the Claude Opus 4.8 model achieved both a high accuracy rate and high agreement values compared to other current models tested. Other analyses comparing various models in radiology cases indicate that Claude-series models sometimes produce more consistent responses when interpreting clinical event sequences than ChatGPT and Gemini derivatives (13). This difference may be related to how intensively the models are exposed to the semantic structure contained in medical journals, case reports, or guidelines such as ACR during their initial training (8). Medical reasoning processes are not simply about concatenating words; the model needs to capture the most critical detail in the patient's clinical presentation and assign it to the correct guideline category. In our study, the Claude Opus 4.8 model appears to perform this synthesis in a more refined manner.

Another finding of our study concerns how a clinical scenario is presented to the AI. Adding a single clinical-scenario example to the question stem significantly improved the models' overall performance and shifted the level of fit from the "significant" range to the "nearly perfect" range. The generalizing nature inherent in LLMs can sometimes lead to models producing overly creative or vague medical advice (11). Just as the human mind understands the framework by looking at a solved example before executing a complex instruction, language models similarly narrow their large vocabulary in response to a single presented example and align themselves with the desired triple "ACR category format." Notably, Claude Opus 4.8 demonstrated high accuracy without disrupting its structure, even when presented without an example. This suggests that the medical question-answer logic may already be preserved within the model's internal parameters, albeit within very distinct limits. As highlighted in the research of Yao et al. (14), it is possible to improve the clinical fit of models from the level of ordinary physicians to the level of experts by applying the correct prompt to the models or providing them with a structural context (8).

Another outcome of our study is that, when viewed through the lens of the most fundamental rule of clinical practice, "first, do no harm," the frequency of models producing dangerous and erroneous decisions remained in the single-digit percentages across all evaluated models. The fact that errors with the potential to harm patients (for example, stating that an essential imaging procedure is unnecessary or recommending a test that should not be performed) constituted a very small minority of the 240 cases examined proves that models can be reliable assistants. An AI capable of identifying the "Usually Not Appropriate" category with 100% accuracy could act as a safety filter in emergency departments, preventing unnecessary tests that might be

overlooked by busy and fatigued physicians (11). However, the decline in the performance of language models in cases where diagnostic boundaries are ambiguous and physicians face the greatest uncertainty, for example right lower quadrant pain that could indicate appendicitis or small bowel obstruction, indicates that AI has not yet fully achieved the delicate balance that an experienced clinician achieves by combining physical examination findings and intuition. Therefore, AI should be positioned not as an authority that replaces the physician but as a consultant that illuminates their blind spots.

An additional consideration is automation bias, the tendency of a busy clinician to accept a system's recommendation without independent verification, particularly under time pressure. If an LLM incorrectly labels a necessary test as "Usually Not Appropriate," and this recommendation is accepted uncritically, the result is not merely a statistical error, but a missed diagnosis with direct medical and legal consequences; responsibility for the final imaging decision and for any resulting harm remains with the ordering physician regardless of the AI input. This underscores that, at the current stage of development, LLM output should be positioned as one input among several, to be reviewed by a clinician before it changes patient management, rather than as an autonomous gatekeeper for imaging decisions.

Regarding the strengths of our study, the latest generation of LLMs was compared impartially using direct, simultaneous, and paired data. To support the clinical relevance of the findings, the analysis was expanded to include not only with simple accuracy percentages but also with weighted kappa statistics, which numerically correct for the severity of errors. The ability to explain how demand strategies affect the model through clinical decision-making mechanisms distinguishes this study from fundamental research in the literature.

Study Limitations

In addition to these strengths, the study has some limitations. The most significant limitation is the small number of cases tested. The small number of scenarios reduced statistical power in the subgroup analyses, preventing clear differentiation of performance between models. The second major limitation is that the clinical cases used were specially prepared, cleaned, and noise-free "ideal" scenarios for training purposes. In real-world hospital settings, physician notes in patient information systems are often disorganized, contain many abbreviations, are incomplete, or contain spelling errors. It would not be accurate to assume that the success of models in these organized, idealized texts will be identical when applied to real-life data. Because the vignettes were constructed to closely mirror the illustrative scenarios in the ACR documents, the scenario text may share phrasing patterns with material to which the models were plausibly exposed during training. This could inflate model performance relative to genuinely novel, real-world clinical narratives. Independent validation using clinician-authored or retrospective cases is an important next step. This focuses only on a small, specific branch of medicine. Since ACR criteria may involve more complex clinical specialties such as traumatology, oncologic follow-up, or neuroradiology, the success levels of the models in those scenarios are unknown.

In light of these considerations, it is essential to take several steps in future clinical and academic studies. First, to measure the true clinical performance of the models, large-scale studies should be designed using real retrospective electronic health records while ensuring patient confidentiality, rather than relying on synthetic cases. Second, analyses should be expanded to include thousands of cases, rather than hundreds, thereby increasing statistical power to cover all sub-branches of radiology. Furthermore, the performance of generic models should be compared with specific models that have been fine-tuned directly with medical data or that have their information confirmed by instantly connecting to external guidelines using techniques such as "Retrieval-Augmented Generation." Beyond fine-tuning or retrieval augmentation, future systems could be designed to reference multiple evidence sources simultaneously, for example, by cross-checking a recommendation against both the ACR Appropriateness Criteria and other relevant guideline bodies, and to present the supporting evidence profile alongside the recommended category, rather than providing a single categorical answer. Such transparency would allow the ordering physician to see not just what the model recommends, but why, which is likely to be a prerequisite for clinical trust and adoption. Finally, the integration of multimodal systems—where language models can analyze both text and a patient's previous X-ray or ultrasound, in alignment with ACR guidelines— should be the next logical focus of medical technologies.

Conclusion

Our study shows that modern LLMs can accurately align radiological imaging recommendations for complex acute abdominal clinical conditions with the scientific standards set by the ACR. Demonstrating the solution to a sample problem in guideline format, rather than simply giving instructions, increased model performance. By helping physicians reach a guideline-concordant imaging decision more quickly, such support has the potential to shorten decision times in the emergency department and reduce patients' exposure to unnecessary ionizing radiation. However, algorithms have not yet reached the level of clinicians'

experience in complex disease presentations, and more extensive research using real-world hospital automation data is needed.

Ethics

Ethics Committee Approval: This study was completed using current versions of publicly available large language models, drawing on ACR's publicly available guidelines. Since no patient data, hospital records, or images were used, ethical committee approval was not required.

Informed Consent: This is cross-sectional study.

Data availability: The prompt texts and clinical scenarios used in the study are provided as supplementary material.

Footnotes

Authorship Contributions

Concept: A.E.A., Design: A.E.A., H.İ.Ş., Data Collection or Processing: H.İ.Ş., F.C.T., Analysis or Interpretation: A.E.A., F.C.T., Literature Search: H.İ.Ş., Writing: A.E.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

Supplementary Material: https://d2v96fxpocvxx.cloudfront.net/72b4fafb-5db4-4d61-bedf-c8100360bb15/content-images/6633bd7b-5abf-47ef-90f1-aae68082fdb9.pdf

References

1. Scheirey CD, Fowler KJ, Therrien JA, Kim DH, Al-Refaie WB, Camacho MA, et al. ACR Appropriateness Criteria(®) acute nonlocalized abdominal pain. J Am Coll Radiol. 2018;15:S217-31.
2. Cartwright SL, Knudson MP. Diagnostic imaging of acute abdominal pain in adults. Am Fam Physician. 2015;91:4529.
3. Russo GK, Zaheer A, Kamel IR, Porter KK, Archer-Arroyo K, Bashir MR, et al. ACR Appropriateness Criteria® Right Upper Quadrant Pain: 2022 Update. J Am Coll Radiol. 2023;20:S211-23.
4. Üçdal M, Ekingen E. Multidisciplinary AI-assisted diagnosis and management of acute abdominal pain syndrome: a pilot study. Eurasian J Emerg Med. 2026;25:272-4.
5. Chang KJ, Marin D, Kim DH, Fowler KJ, Camacho MA, Cash BD, et al. ACR Appropriateness Criteria® Suspected Small-Bowel Obstruction. J Am Coll Radiol. 2020;17:S305-14.
6. Kambadakone AR, Santillan CS, Kim DH, Fowler KJ, Birkholz JH, Camacho MA, et al. ACR Appropriateness Criteria® Right Lower Quadrant Pain: 2022 Update. J Am Coll Radiol. 2022;19:S445-61.
7. Pambudi S, Menolascina F. Bridging clinical narratives and ACR appropriateness guidelines: a multi-agent RAG system for medical imaging decisions. arXiv. 2025.
8. Papageorgiou PS, Christodoulou RC, Pitsillos R, Petrou V, Vamvouras G, Kormentza EV, et al. The role of large language models in improving diagnostic-related groups assignment and clinical decision support in healthcare systems: an example from radiology and nuclear medicine. Appl Sci. 2025;15:9005.
9. Rau A, Rau S, Zoeller D, Fink A, Tran H, Wilpert C, et al. A Context-based chatbot surpasses trained radiologists and generic ChatGPT in following the ACR appropriateness guidelines. Radiology. 2023;308:e230970.
10. Zaki HA, Aoun A, Munshi S, Abdel-Megid H, Nazario-Johnson L, Ahn SH. The application of large language models for radiologic decision making. J Am Coll Radiol. 202;21:1072-8.
11. Tziakouri A, Menolascina F. Reinforcement learning for clinical reasoning: aligning LLMs with ACR imaging appropriateness criteria. arXiv [Preprint]. 2025;arXiv:2510.05194.
12. American College of Radiology. ACR Appropriateness Criteria® [Internet]. [cited 2026 Jul 1]. Available from: https://acsearch.acr.org/list.
13. Sonoda Y, Kurokawa R, Nakamura Y, Kanzawa J, Kurokawa M, Ohizumi Y, et al. Diagnostic performances of GPT-4o, Claude 3 Opus, and Gemini 1.5 Pro in "diagnosis please" cases. Jpn J Radiol. 2024;42:1231-5.
14. Yao MS, Chae A, Saraiya P, Kahn CE Jr, Witschey WR, Gee JC, et al. Evaluating acute image ordering for real-world patient cases via language model alignment with radiological guidelines. Commun Med (Lond). 2025;5:332.

EURASIAN JOURNAL OF EMERGENCY MEDICINE

DOI: 10.4274/eajem.galenos.2026.e002

Erratum

Eurasian J Emerg Med. 2026;25: 419

DOI: 10.4274/eajem.galenos.2026.59752

In the published article, the affiliation of the author Durmuş Ayan was inadvertently stated incorrectly on the first page. The affiliation has been corrected to reflect the author's correct institutional information.

Incorrect affiliation of Durmuş Ayan (affiliation 3):

Niğde Ömer Halisdemir University Faculty of Medicine, Niğde Ömer Halisdemir University Training and Research Hospital, Department of Emergency Medicine, Niğde, Türkiye

Correct affiliation of Durmuş Ayan (affiliation 3):

Niğde Ömer Halisdemir University Faculty of Medicine, Niğde Ömer Halisdemir University Training and Research Hospital, Department of Medical Biochemistry, Niğde, Türkiye

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.24445
Eurasian J Emerg Med. 2026;25: 420-8

Heat-Waves as a Growing Global Health Crisis: Multi-decadal Evidence from 1936 to 2024

Mustafa Öcal

Ufuk University Dr. Rıdvan Ege Hospital, Department of Emergency Medicine, Ankara, Türkiye

Abstract

Aim: Heat-wave disasters are among the most lethal and rapidly escalating categories of natural hazards. Despite growing evidence of their increasing frequency and mortality burden, a comprehensive multi-decadal global assessment has been lacking.

Materials and Methods: Using Emergency Events Database records from 1936 to 2024, we quantified heat-wave disaster frequency, mortality, and geographic distribution. Linear regression was applied to assess temporal trends at annual and decadal scales, and the Kruskal-Wallis test with Dunn post-hoc analysis was used to compare per-event mortality across continents.

Results: A total of 344 heat-wave disasters were recorded across 75 countries, resulting in 350,425 deaths. Europe recorded the highest number of events (n=184; 53.5%) and the overwhelming majority of fatalities (320,418 deaths; 91.4%), despite comprising predominantly high-income nations—a paradox driven by aging populations, low air-conditioning penetration, and urban heat island effects. India experienced the highest event frequency (n=32), while the Russian Federation recorded the highest mean deaths per disaster (18,677.3; n=3 events). Disaster frequency increased significantly over time (β=0.170 events year $^{-}$(1); $p<0.001$; $R(2)=0.294$), as did annual mortality (β=204.1 deaths year $^{-}$(1); $p<0.001$). Decadal aggregation revealed even stronger relationships ($R(2)=0.725$ and 0.592, respectively). May through August accounted for 87.8% of events and 98.4% of global deaths.

Conclusion: Heat-wave disasters are intensifying in frequency and lethality globally. Urgent investment in early warning systems, equitable cooling infrastructure, and evidence-based climate adaptation policy is essential to reduce future heat-wave mortality.

Keywords: Heat-wave disaster, climate change, extreme temperature, vulnerability, Emergency Events Database, global disaster epidemiology

Introduction

Heat-waves-sustained periods of anomalously high temperatures relative to local climatological norms-rank among the most lethal natural hazards on Earth. Unlike floods or storms, their destruction is largely invisible: heat causes death through a cascade of physiological failures including heat stroke, cardiovascular decompensation, acute kidney injury, and respiratory compromise, disproportionately affecting the elderly, outdoor workers, and individuals with chronic disease (1,2). Globally, an estimated 489,000 heat-related deaths occurred annually between 2000 and 2019 (1-3), and this burden is accelerating. 2023 was the hottest year on record, with approximately 178,486 excess deaths attributed to heat-waves worldwide, more than half of which were directly attributable to human-induced climate change (3).

The true scale of heat-wave mortality was brought to global attention by the European summer of 2003, which caused more than 70,000 excess deaths across the continent (4) a single event that redefined heat as a major public health emergency. The 2010 Russian heat-wave caused approximately 56,000 deaths in 44 days (5), and subsequent European summers in 2022 and in 2023 resulted in an estimated 61,672 (6) and 47,690 heat-related deaths (7). These repeated catastrophes reflect a clear physical reality: the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) concluded with virtual certainty that the frequency and intensity of hot extremes have increased

Corresponding Author: Mustafa Öcal MD, Ufuk University Dr. Rıdvan Ege Hospital, Department of Emergency Medicine, Ankara, Türkiye
E-mail: mustafaocal2244@gmail.com **ORCID ID:** orcid.org/0000-0001-8432-9695
Cite this article as: Öcal M. Heat-waves as a growing global health crisis: multi-decadal evidence from 1936 to 2024. Eurasian J Emerg Med. 2026;25: 420-8

Received: 02.05.2026
Accepted: 01.09.2026
Published: 22.09.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

globally since the 1950s, driven primarily by anthropogenic greenhouse gas emissions, and they will continue to intensify under all future warming scenarios (8). At 2 °C of global warming, heat-mortality events that historically recurred once per century are projected to occur at least once every five years (9).

Despite this escalating burden, the global multi-decadal epidemiology of heat-wave disasters-as formally recorded events meeting standardized impact thresholds-has not been comprehensively characterized. Prior research has predominantly focused on temperature-mortality associations in specific cities, on excess mortality during individual events, or on modeled projections of future heat burden. Critically, no study has leveraged the Emergency Events Database (EM-DAT), the world's most comprehensive standardized disaster registry (10), to analyze the full historical record of heat-wave disasters across all continents and income groups over a multi-decadal horizon. As a result, the long-term evolution of heat-wave disaster frequency, mortality burden, and geographic vulnerability patterns remains systematically uncharacterized (11).

This study addresses these critical gaps by analyzing all heat-wave disasters recorded in EM-DAT from 1936 to 2024-the most temporally comprehensive global assessment of its kind. We quantify temporal trends in disaster frequency and mortality at annual and decadal scales, describe the geographic distribution of events and deaths across continents and countries, and evaluate vulnerability using deaths per disaster as a proxy for adaptive capacity. The findings provide a rigorous, evidence-based global perspective on the intensifying burden of heat-wave disasters and offer actionable insights for climate adaptation and disaster risk reduction policy.

Materials and Methods

Data Source

All data were obtained from the (EM-DAT; www.emdat.be), maintained by the Centre for Research on the Epidemiology of Disasters at the Université Catholique de Louvain, Belgium. EM-DAT compiles standardized information on natural and technological disasters from national authorities, United Nations agencies, non-governmental organizations, and peer-reviewed literature, applying multi-source validation and internal quality-control procedures before events are published (10).

Records categorized under the "heat-wave" disaster subtype within the "extreme temperature" disaster type were extracted for the full available period (1936–2024). EM-DAT applies the following predefined inclusion criteria for disaster registration: (1) ≥10 deaths; (2) ≥100 people affected; (3) declaration of a state of emergency; or (4) a request for international assistance. All records meeting at least one of these criteria were included without additional filtering. Events with missing temporal or geographic information were excluded. The absence or low frequency of recorded events in the early decades (particularly between 1936 and 1950) likely reflects limitations in global disaster reporting systems rather than indicating a true absence of heat-wave events. The dataset was accessed in April 2026 (version 2026-04-09) (11,12). This study utilized publicly available, de-identified secondary data from the EM-DAT. The dataset does not include any personal, identifiable, or sensitive human subject information. Therefore, according to institutional policies and international research ethics guidelines, formal ethical approval and informed consent were not required. The study was conducted in accordance with the Declaration of Helsinki and relevant research standards.

Operational Definitions and Derived Variables

A heat-wave disaster was operationally defined as a sustained period of anomalously high temperature resulting in human casualties or economic losses sufficient to meet EM-DAT inclusion criteria. For each event, EM-DAT provided the year of occurrence, affected country, continent/region, number of deaths, and reported economic damage in USD. Annual and decadal totals were generated directly from the reported event year. Decades were defined as ten-year intervals commencing with the year ending in zero (e.g., 1930–1939, 1950–1959).

"Deaths per disaster" was calculated by dividing total recorded deaths by total events for each country, serving as a proxy indicator of disaster vulnerability and national adaptive capacity. Death data were available for 314 of 344 events (91.3%); events lacking death records were retained in frequency analyses but excluded from mortality-specific calculations. Economic loss data were available for only 21 of 344 events (6.1%) and were not subjected to formal statistical analysis.

Statistical Analysis

All statistical analyses were performed using IBM SPSS Statistics (version 29.0; IBM Corp., Armonk, NY, USA). Descriptive statistics-including total event counts, total and mean deaths per-event, median deaths per-event, and percentage distributions-were computed to characterize the global and regional burden of heat-wave disasters. The top 10 countries by event frequency and by total mortality, and the top 10 countries by mean deaths per disaster, were tabulated to assess geographic concentration. Because mean deaths per disaster is sensitive to the number of recorded events underlying each country's estimate, this ranking was additionally re-derived after applying minimum event-count thresholds of ≥3 and ≥5 events per country, as a sensitivity check on the stability of top-ranked countries.

Temporal trend analyses were conducted at two scales. At the annual-scale, simple linear regression was applied to the complete yearly time series (1936–2024), years with no recorded events were assigned a value of zero and were included in the regression, so that non-event years contributed to the trend estimate rather than being excluded. Because 35 of 89 years (39.3%) had zero recorded events, the annual-scale series was zero-skewed and right-skewed. Ordinary least-squares regression on such data can understate or distort the true trend. Therefore, the annual-scale was treated as a descriptive, conservative check rather than the primary inferential estimate, with the decadal-scale analysis serving as the principal trend test (Study Limitations). At the decadal scale, event counts and deaths were aggregated into nine decade-level observations (1930s through 2020s), a and simple linear regression was applied. For each model, the slope ($\beta \pm$ standard error), coefficient of determination (R(2)), and two-tailed p-value were reported. Statistical significance was defined as $\alpha=0.05$. Because the decadal-scale model is based on only nine decade-level observations, its statistical power is inherently limited, and the associated R(2) and p-values should be interpreted as a description of the aggregate trend rather than as evidence with the precision typically implied by a large-sample regression; it is reported alongside and is concordant in direction with the annual-scale model, and both are considered jointly rather than in isolation (Study Limitations). To evaluate whether the observed trends were artifacts of improved reporting completeness following EM-DAT's adoption of standardized inclusion criteria in the 1990s, a sensitivity analysis was conducted by repeating the annual-scale regressions on the 1990–2024 subset.

Seasonal distribution was assessed by tabulating events and deaths by month of disaster onset. Continental comparisons of per-event mortality were conducted as follows: First, the normality of the per-event mortality distributions was assessed using the Shapiro-Wilk test. All continental groups showed significant deviations from normality ($p<0.05$). The global distribution exhibited marked positive skewness (skewness =8.84; kurtosis =101.26), indicating a highly right-skewed, heavy-tailed distribution. Accordingly, comparisons were performed using the Kruskal-Wallis H test, followed by Dunn's post-hoc pairwise analysis with Bonferroni correction. P-values were adjusted using the Bonferroni correction. Statistical significance was defined as follows: $^{*}p<0.05$, $^{**}p<0.01$, and $^{*}p<0.001$ (ns: not significant).

Data Availability

All data are publicly accessible from EM-DAT (www.emdat.be). Derived datasets and code are available from the corresponding author upon reasonable request.

Results

Between 1936 and 2024, 344 heat-wave disasters were recorded in 75 countries across five continents. The earliest recorded event dates to 1936; no events were recorded between 1937 and 1950, which is consistent with the early-period reporting limitations of international disaster databases. Death data were available for 314 of 344 events (91.3%). In total, these disasters resulted in 350,425 deaths, with a mean of 1,116 deaths per disaster and a median of 132.5 deaths, reflecting a highly right-skewed distribution driven by a small number of catastrophic events. The single most lethal event was the 2010 Russian heat-wave (55,736 deaths).

Europe recorded both the highest number of events (n=184; 53.5% of the global total) and an overwhelming majority of fatalities (320,418 deaths; 91.4% of the global total), with a mean of 1,873.8 deaths per disaster. This striking concentration of mortality in predominantly high-income European nations, despite their relatively advanced health infrastructure, represents a paradox examined in detail in the Discussion. Asia ranked second in event frequency (n=102), but contributed only 5.9% of global mortality (20,777 deaths). The Americas, Africa, and Oceania accounted for 2.4%, 0.1%, and 0.1% of global deaths, respectively (Table 1, Figure 1).

The Kruskal-Wallis test confirmed that per-event mortality differed significantly across continents ($H=29.80$, $p<0.001$). Dunn's post-hoc analysis with Bonferroni correction identified Europe as significantly different from Asia ($p=0.0005$), the Americas ($p=0.036$), and Africa ($p=0.033$); all other pairwise comparisons were non-significant.

At the country level, India experienced the highest number of heat-wave disasters (n=32; 9.3% of the global total), followed by the United States of America (27), Pakistan (19), Japan (17), and France, Romania, and Belgium (11 each) (Figure 2).

In terms of mortality, the ranking differed markedly and was dominated by European nations. Italy recorded the highest total deaths (69,909; 19.9% of global fatalities), followed by the Russian Federation (56,032; 16.0%), Spain (41,597; 11.9%), France (37,509; 10.7%), and Germany (30,192; 8.6%). The top five countries-all in Europe accounted for 67.1% of global heat-wave deaths. India ranked seventh (13,291 deaths; 3.8%) despite leading in event frequency (Figure 2).

The mean number of deaths per disaster-a proxy indicator of disaster vulnerability and adaptive capacity varied markedly across countries. The Russian Federation recorded the highest deaths per-event (18,677.3; based on only n=3 recorded events),

a figure largely driven by the 2010 heat-wave. Italy ranked second (7,767.7; n=9 events), followed by Spain (4,621.9; n=9 events), Germany (3,774.0; n=8 events), and France (3,409.9; n=11 events). All of the top 10 countries by this metric were European or high-income nations, underscoring a pattern that is strikingly distinct from cold-wave disasters, in which vulnerability is concentrated in lower-income countries.

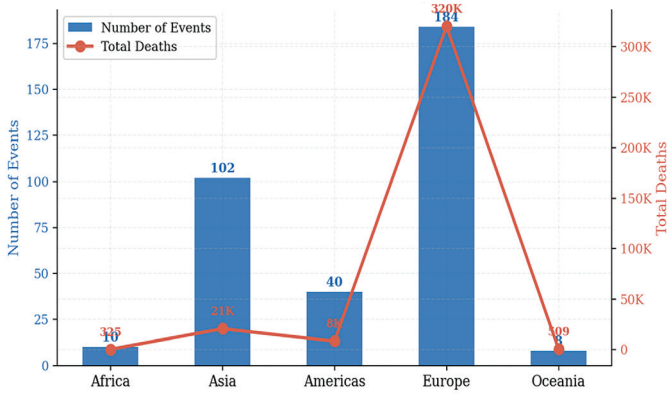


Figure 1. Heat-wave disasters and associated deaths by continent (1936–2024)

Linear regression revealed a statistically significant upward trend in annual heat-wave disaster frequency (β=0.170±0.028 events year $^{-1}$; p<0.001; R(2)=0.294). Years with no recorded events (n=35) were assigned a value of zero and included in the regression, providing a conservative estimate of the long-term trend. Annual mortality also showed a significant upward trend (β=204.1±55.1 deaths year $^{-1}$; p<0.001; R(2)=0.136). The comparatively low R(2) for mortality reflects the highly skewed, non-stationary nature of annual death counts (coefficient of variation =3.63), driven by the disproportionate influence of catastrophic outlier events such as 2003 (73,722 deaths), 2010 (56,448 deaths), 2022, and 2024, each of which exceeded 60,000 deaths. These levels were previously seen only during the most catastrophic single-year events on record (Table 2, Figure 3).

When aggregated by decade, both trends were considerably stronger. Disaster frequency increased from 2 events in the 1930s to 131 events in the 2020s (through 2024), with decadal models explaining 72.5% of variance in event counts (β=1.298±0.302; p=0.004; R(2)=0.725). Mortality rose from 1,693 deaths in the 1930s to 184,128 deaths in the 2020s, with decadal models

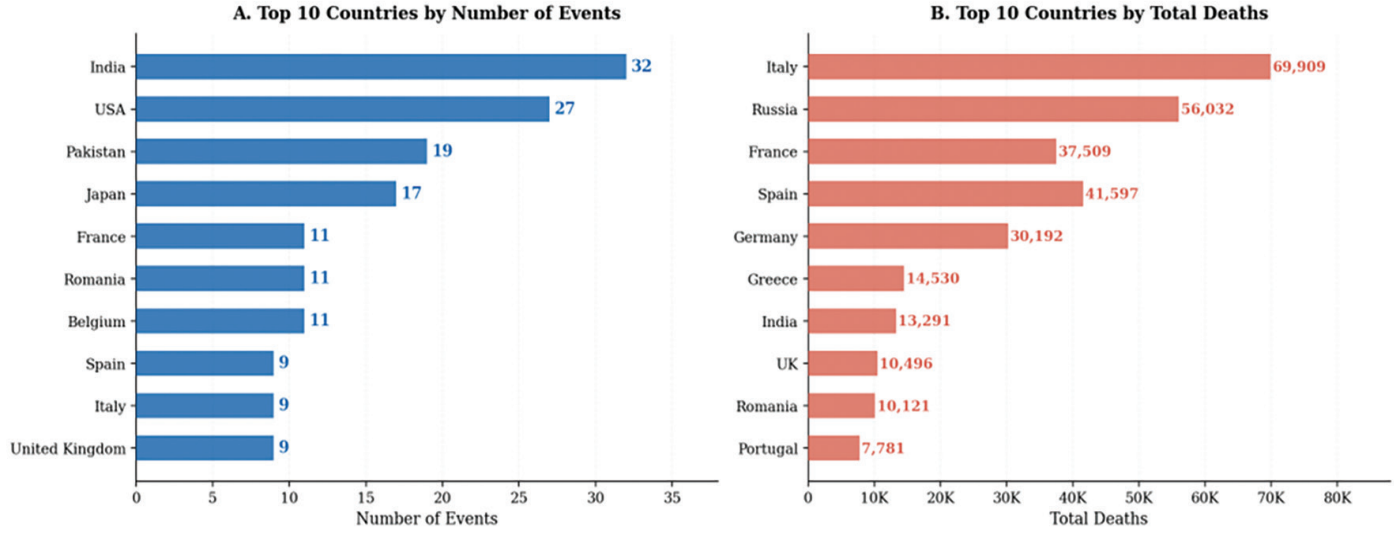


Figure 2. Country-level analysis of heat-wave disasters (1936–2024)

Table 1. Global and continental summary of heat-wave disasters

Continent	No. of disasters	Total deaths	Mean deaths/event	Share of global deaths (%)
Africa	10	325	36.1	0.1%
Asia	102	20,777	233.4	5.9%
Americas	40	8,396	209.9	2.4%
Europe	184	320,418	1,873.8	91.4%
Oceania	8	509	101.8	0.1%
Global total	344	350,425	—	100%

explaining 59.2% of variance in deaths (β=1,658.0±520.0; p=0.015; R(2)=0.592).

Because this metric is highly sensitive to the number of recorded events per country (Study Limitations), a sensitivity analysis was performed by re-ranking countries after applying minimum event-count thresholds of ≥3 and ≥5 events (Table 3). Under a ≥3-event threshold, Saudi Arabia (n=1) was excluded but the remaining ranking was unchanged, since the Russian Federation (n=3) still qualified. Under the more conservative ≥5-event threshold, however, the Russian Federation no longer qualified, and Italy became the top-ranked country (7,767.7 deaths per event), followed by Spain, Germany, France, Greece, Bulgaria, the United Kingdom, Portugal, Romania, and Croatia, all supported by ≥5 recorded events. This indicates that the Russian Federation's top ranking is not a stable, broadly supported vulnerability signature, but is instead driven almost entirely by a single catastrophic event (the 2010 heat-wave) and should be interpreted accordingly.

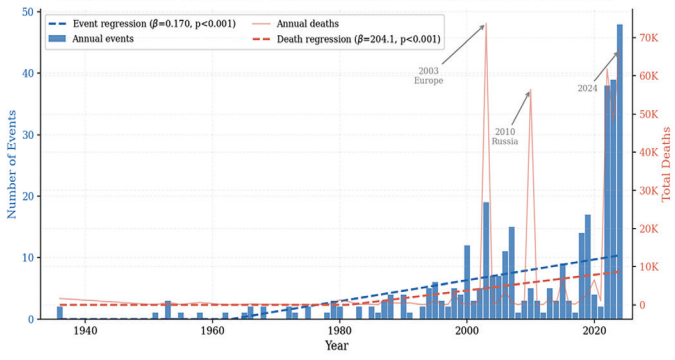


Figure 3. Yearly trends: heat-wave disaster count and total deaths with regression lines (1936–2024)

The sensitivity analysis restricted to 1990–2024 (35 years, of which only 1 had zero recorded events) confirmed both trends: disaster frequency continued to increase significantly (β=0.566±0.166 events year $^{-}$(1); p=0.002; R(2)=0.261), as did mortality (β=798.6±341.1 deaths year $^{-}$(1); p=0.025; R(2)=0.142). Notably, the post-1990 slope of disaster frequency was more than three times as steep as the full-period estimate (0.566 vs. 0.170 events year $^{-}$(1)), indicating that the upward trend is not an artifact of sparse early-period reporting and that, if anything, the full 1936–2024 series understates it.

All 344 events had a recorded start month. Heat-wave disasters exhibited a strongly seasonal pattern concentrated in the Northern Hemisphere summer months. July recorded the highest number of events (n=133; 38.7%), followed by June (n=86; 25.0%) and May (n=57; 16.6%). Collectively, May through August accounted for 87.8% of all events and 98.4% of global deaths. Notably, June recorded the highest mortality (135,717

Table 2. Trend analysis of global heat-wave disasters

Parameter	Time scale	β (slope ± SE)	R(2)	p value
Disaster counts vs. year	Annual	0.170 ± 0.028	0.294	<0.001
Death count vs. year	Annual	204.1 ± 55.1	0.136	<0.001
Disaster counts vs. decade	Decadal	1.298 ± 0.302	0.725	0.004
Death count vs. decade	Decadal	1,658.0 ± 520.0	0.592	0.015

β: Regression slope, SE: Standard error, R(2): Coefficient of determination. All four associations were statistically significant (p<0.05)

Table 3. Sensitivity of the deaths-per-disaster ranking to minimum event-count thresholds

Country	n events	Mean deaths/event	≥3 events	≥5 events
Russian Federation	3	18,677.3	Yes	No
Italy	9	7,767.7	Yes	Yes
Spain	9	4,621.9	Yes	Yes
Germany	8	3,774.0	Yes	Yes
France	11	3,409.9	Yes	Yes
Greece	8	1,816.2	Yes	Yes
Saudi Arabia	1	1,301.0	No	No
Bulgaria	5	1,274.0	Yes	Yes
United Kingdom	9	1,166.2	Yes	Yes
Portugal	7	1,111.6	Yes	Yes

The countries shown are the unfiltered top 10 by mean deaths per disaster (n = total recorded events per country). Shaded rows are excluded from the ranking under a ≥ 5-event threshold. Under ≥5, the ranking becomes: Italy, Spain, Germany, France, Greece, Bulgaria, United Kingdom, Portugal, Romania (920.1), and Croatia (675.0)

deaths; 38.7%), exceeding that of July (88,684; 25.3%), reflecting the disproportionate impact of the 2003 European heat-wave, which peaked in June and in early August. No events were recorded in September or October across the entire study period; this may reflect both a lower frequency of meteorological heat-waves in post-summer months and potential underreporting or categorical-boundary effects that cause late-season events to fail to meet EM-DAT inclusion thresholds or historical subtype coding conventions (Figure 4).

Discussion

This study provides the first comprehensive, multi-decadal assessment of heat-wave disasters on a global scale using EM-DAT records spanning 1936–2024. A total of 344 disasters were recorded across 75 countries, resulting in 350,425 deaths. Both annual and decadal regression analyses confirmed statistically significant upward trends in disaster frequency and mortality, with the 2020s emerging as the deadliest decade on record (131 events; 184,128 deaths) a trajectory consistent with the accelerating warming documented in IPCC AR6 (8). Europe dominated the global burden, accounting for 91.4% of all heat-wave deaths despite representing 53.5% of recorded events. The Russian Federation recorded the highest mean deaths per disaster (18,677.3; n=3 events; see sensitivity analysis below), followed by Italy (7,767.7) and Spain (4,621.9), while India led in event frequency (32 disasters). The strong seasonal concentration of events in May–August (87.8% of events; 98.4% of deaths), with June recording the highest mortality despite July having the highest event count, reflects both the climatological drivers of Northern Hemisphere heat-waves and the disproportionate impact of the 2003 European event. Together, these findings underscore that heat-waves are not merely an environmental phenomenon but a rapidly intensifying public health crisis demanding urgent, globally coordinated action.

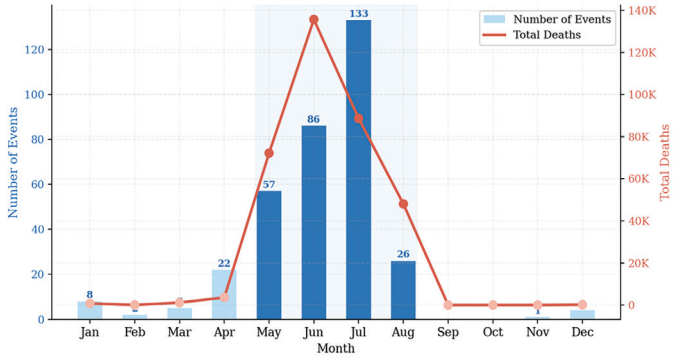


Figure 4. Monthly distribution of heat-wave disasters and associated deaths (1936–2024)

The observed increase in annual heat-wave disaster frequency (β=0.170 events year ⁻(1); R(2)=0.294) and the dramatic decadal acceleration-from 6 events in the 1950s to 131 in the 2020s are consistent with the robust physical attribution of heat extreme intensification to anthropogenic greenhouse gas emissions (8). At 2 °C of global warming, heat-mortality events that occurred once per century in the early 2000s are projected to occur at least once every five years (9). The particularly sharp rise in the 2020s reflects the confluence of record-breaking temperatures and growing human exposure through rapid urbanization and population growth. In 2023 alone, approximately 178,486 excess deaths were globally attributed to heat-waves, of which more than 54% were attributable to human-induced climate change (3), suggesting that EM-DAT-based disaster counts represent only the visible apex of a far larger heat-mortality burden.

One of the most striking findings is the extraordinary concentration of heat-wave mortality in Europe 91.4% of global deaths despite being composed predominantly of high-income nations with developed health infrastructure. This apparent paradox can be understood through several interconnected mechanisms. First, European populations are less physiologically and infrastructurally adapted to extreme heat than populations in tropical regions, with residential air-conditioning penetration remaining substantially lower in Northern and Central Europe (13). Second, Europe has one of the world's most rapidly aging populations: heat-related mortality rates are 768% higher among individuals aged 80 years and older than those aged 65–79 years (7). Third, rapid urbanization and the urban heat island effect amplify background temperatures within European cities, elevating local temperatures several degrees above surrounding rural areas and increasing population exposure (13). Fourth, the 2003 European heat-wave exposed critical systemic failures in early warning, social care coordination, and hospital surge capacity, producing more than 70,000 excess deaths (4).

Since 2003, European countries have progressively implemented national heat health action plans. Modeling suggests that without present-century societal adaptation, the 2023 European heat-related mortality burden would have been 80% higher (7). Nevertheless, the persistence of tens of thousands of heat-related deaths annually with an estimated 62,775 in Europe in 2024 alone (14) demonstrates that adaptation remains insufficient relative to the pace of warming. Projections indicate heat-related mortality in Europe is expected to increase by 103.7–135.1 deaths per million people per degree of global warming by 2100 under various emissions scenarios, with hotspots concentrated in Southern Europe and increasingly in Northern and Eastern Europe due to demographic aging (15).

The Russian Federation recorded the highest mean deaths per disaster (18,677.3), driven primarily by the catastrophic 2010 heat-wave a 44-day event that caused an estimated 55,736 deaths and was the deadliest single heat-wave disaster in the EM-DAT record. Several compounding factors explain this extreme lethality. Meteorologically, the 2010 event was unprecedented in both intensity and duration for Russia, with daytime temperatures reaching 38.2 °C in Moscow (16) values far outside the historical range to which infrastructure and populations were acclimatized. At the institutional level, Russia lacked a formalized national heat health action plan, early-warning dissemination was inadequate, and the public health response was delayed relative to the pace of the disaster (5). This case underscores that even in countries not typically classified as heat-vulnerable, the absence of preparedness systems can transform a meteorological extreme into a public health catastrophe. This ranking should, however, be interpreted with caution: it rests on only three recorded events, one of which, the 2010 disaster, accounts for 99.5% of the country's cumulative heat-wave deaths. When a minimum event-count threshold of ≥5 is applied as a sensitivity check (Results; Table 3), the Russian Federation no longer meets the threshold and Italy becomes the top-ranked country. The 2010 Russian heat-wave therefore remains a genuine and instructive catastrophic outlier, but the "highest mean deaths per disaster" ranking itself is a statistically fragile, single-event-driven statistic rather than evidence of a stable national vulnerability profile.

Asia contributed 102 events (29.7% of the global total) but accounted for only 5.9% of recorded deaths, with India alone accounting for 32 disasters. High event frequency in South Asia is driven by structural factors including dense agricultural labor forces with prolonged outdoor exposure, pervasive energy poverty limiting cooling access, rapid urbanization without adequate heat island mitigation, and health systems with limited surge capacity (17). Climate change has rendered South Asian heat-waves approximately 45 times more likely than in a pre-industrial climate (18), and wet-bulb temperatures in the Indus and Ganges River valleys are projected to approach or exceed the physiological survivability threshold of 35 °C under high-emission scenarios (19). The lower per-event mortality relative to Europe does not indicate lower vulnerability; rather, it reflects that many events fail to reach the catastrophic single-event threshold seen in European heat-waves, while chronic heat-mortality remains substantially underreported (17). India, Pakistan, and Bangladesh have taken initial steps toward heat action planning-most notably Ahmedabad's pioneering framework following the 2010 heat-wave (17) but implementation remains geographically uneven and insufficiently scaled relative to the magnitude of the risk (20).

The strong concentration of heat-wave disasters in May–August reflects the dominant role of Northern Hemisphere summer meteorology in shaping global patterns. The observation that June recorded the highest mortality (135,717 deaths; 38.7%) despite July having the highest event count is primarily attributable to the 2003 European heat-wave, which peaked in June and early August and produced catastrophic simultaneous mortality across multiple countries (4). This highlights that peak disaster impact does not necessarily coincide with peak event frequency, and that the epidemiological burden is disproportionately shaped by a small number of catastrophic outlier events. The complete absence of recorded events in September and October warrants careful interpretation: although Northern Hemisphere heat-wave frequency declines sharply after August, a reporting artifact cannot be excluded because late-season events may produce mortality below EM-DAT's inclusion thresholds. A distinct but related possibility is definitional rather than purely statistical: if earlier EM-DAT coding practice or contributing national sources implicitly bounded the "heat-wave" category to the conventional Northern Hemisphere summer window, late-season warm spells occurring in September or October may have been classified under a different hazard category (or omitted) rather being genuinely absent; this would produce the same zero count without reflecting a true climatological absence of events. Future research integrating EM-DAT records with meteorological reanalysis data (e.g., ERA5) would help disentangle genuine seasonal absence from potential underreporting or categorical-boundary effects.

Policy and Adaptation Implications

Heat-waves must be prioritized as a primary focus of climate adaptation. Early warning systems represent the highest-yield near-term intervention for reducing heat-wave mortality. France's national heat health action plan, implemented in 2004 following the catastrophic 2003 event, demonstrated measurable effectiveness during the 2006 heat-wave, where observed mortality (2,065 deaths) was substantially lower than the predicted toll of 6,452 deaths had no plan been in place suggesting more than 4,000 deaths averted by the intervention (21). Japan's integrated heat stroke prevention framework and Australia's state-level heat action plans similarly demonstrate that structured early-warning governance can substantially reduce per-event lethality (18). However, a global assessment found that cities in high- and upper-middle-income countries described full early warning systems nearly three times more often than those in lower-income countries (22), highlighting a critical equity gap that must be addressed through international technology transfer and climate finance, consistent with the Sendai Framework for Disaster Risk Reduction 2015–2030 (23).

Investment in passive cooling infrastructure-such as urban green spaces, reflective building surfaces, natural ventilation design, and access to affordable air-conditioning-is necessary even in high-income settings. For South Asia and sub-Saharan Africa, expanding access to cooling while reducing energy poverty represents the most critical near-term intervention pathway. Heat-wave risk must be mainstreamed into national adaptation plans under the Paris Agreement and integrated with urban planning, healthcare system preparedness, and social protection frameworks.

Study Limitations

The principal strength of this study is its comprehensive temporal coverage (1936–2024) and its use of a globally standardized, publicly accessible disaster database, enabling consistent comparison across decades and continents (10). Several limitations warrant consideration. EM-DAT relies on secondary reporting, and underreporting is a recognized limitation, particularly in low-income regions. Indirect heat-related deaths are substantially underreported, indicating that the true mortality burden far exceeds EM-DAT estimates. Economic damage data were available for only 21 of 344 events (6.1%), precluding assessment of the financial burden. The database also lacks meteorological parameters-such as maximum temperature, duration, and humidity-that would allow direct linkage between physical hazard intensity and health outcomes. Notably, the deaths per disaster metric is sensitive to the number of recorded events per country: countries with a single recorded event (e.g., Saudi Arabia, n=1) will have their mean entirely determined by that event, potentially inflating or deflating the metric relative to countries with multiple events. This limitation should be considered when interpreting cross-country comparisons; it is addressed quantitatively via a minimum event-count threshold sensitivity analysis (Table 3).

An important caveat is the improved observational coverage and reporting accuracy since the 1990s, when EM-DAT adopted standardized inclusion criteria. This likely contributes significantly to the observed increase in recorded disaster frequency, particularly in earlier decades and should be explicitly considered when interpreting long-term trends. To address this, a sensitivity analysis restricted to post-1990 data-the period of more consistent reporting-was conducted and confirmed both trends, indicating that the increases in disaster frequency and mortality are not solely artifacts of underreporting in earlier periods. Two further statistical limitations should be noted. First, 39.3% of years (35/89) in the annual-scale series had zero recorded events; this zero-inflation and marked right-skewness mean that ordinary least-squares regression may not fully capture the underlying trend, and count-based models (e.g., Poisson or negative binomial regression) would be a more statistically appropriate approach for future analyses of this series. Second, the decadal-scale regression relies on only nine aggregated observations; estimates of R(2) and p from so few points have limited statistical power and should be regarded as descriptive of the aggregate trend rather than high-precision inferential estimates. We present it as a corroborating, not stand-alone, analysis alongside the annual-scale model. Finally, EM-DAT does not record age- or sex-disaggregated mortality data, which precludes the calculation of age-adjusted mortality rates or sex-stratified analyses. Both approaches are near-standard in heat-mortality epidemiology and would have allowed direct comparison with demographic-adjustment studies (7,13). This represents an important gap that could be addressed by linking EM-DAT records to national vital statistics or WHO mortality databases in future work. Future research should integrate EM-DAT records with gridded climate reanalysis products and incorporate socioeconomic indicators-including air-conditioning penetration, age structure, urban heat island intensity, and health-system capacity-into multivariate vulnerability models to enable more granular global risk assessment (24).

Conclusion

Between 1936 and 2024, 344 heat-wave disasters caused 350,425 deaths across 75 countries, with both the frequency of disasters and mortality increasing significantly over the period. The 2020s have already surpassed all previous decades in the number of events and deaths, reflecting the accelerating impact of anthropogenic climate change. Europe's disproportionate burden-91.4% of global deaths-demonstrates that economic development alone does not confer heat resilience, while Asia's high event frequency underscores the distinct vulnerability of densely populated, rapidly urbanizing regions with limited adaptive capacity.

These findings underscore the need for urgent action. Heat-waves are increasing in frequency, intensity, and lethality. Strengthening early warning systems, expanding equitable access to cooling infrastructure, and mainstreaming heat risk into national adaptation frameworks and urban planning are essential priorities. International cooperation and climate finance directed toward the most vulnerable populations are indispensable if the goals of the Sendai Framework, the Paris Agreement, and Sustainable Development Goal 13 are to be achieved. Heat-waves are already among the deadliest natural hazards globally, and without rapid and coordinated adaptation, their impact is expected to escalate beyond the capacity of current health and disaster response systems.

Ethics

Ethics Committee Approval: This study utilized publicly available, de-identified secondary data from the EM-DAT (Emergency Events Database). The dataset does not include any personal, identifiable, or sensitive human subject information. Therefore, according to institutional policies and international research ethics guidelines, formal ethical approval and informed consent were not required. The study was conducted in accordance with the Declaration of Helsinki and relevant research standards.

Informed Consent: Not applicable. The study was based exclusively on publicly available, de-identified secondary data and did not involve direct participation of human subjects; therefore, informed consent was not required.

Footnotes

Conflict of Interest: No conflict of interest was declared by the author.

Financial Disclosure: The author declared that this study received no financial support.

References

1. Arsad FS, Hod R, Ahmad N, Ismail R, Mohamed N, Baharom M, et al. The impact of heatwaves on mortality and morbidity and the associated vulnerability factors: a systematic review. Int J Environ Res Public Health. 2022;19:16356.
2. Gasparrini A, Guo Y, Hashizume M, Lavigne E, Zanobetti A, Schwartz J, et al. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. Lancet. 2015;386:369-75.
3. Hundessa S, Huang W, Xu R, Yang Z, Zhao Q, Gasparrini A, et al. Global excess deaths associated with heatwaves in 2023 and the contribution of human-induced climate change. Innovation (Camb). 2025;6:101110.
4. Robine JM, Cheung SLK, Le Roy S, Van Oyen H, Griffiths C, Michel JP, et al. Death toll exceeded 70,000 in Europe during the summer of 2003. C R Biol. 2008;331:171-8.
5. World Health Organization. Heat and Health. WHO Fact Sheet. Geneva, Switzerland: WHO; 2024.
6. Ballester J, Quijal-Zamorano M, Méndez Turrubiates RF, Pegenaute F, Herrmann FR, Robine JM, et al. Heat-related mortality in Europe during the summer of 2022. Nat Med. 2023;29:1857-66.
7. Gallo E, Quijal-Zamorano M, Méndez Turrubiates RF, Tonne C, Basagaña X, Achebak H, et al. Heat-related mortality in Europe during 2023 and the role of adaptation in protecting health. Nat Med. 2024;30:3101-05.
8. Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, et al., editors. Climate change 2021: the physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press; 2021.
9. Lüthi S, Fairless C, Fischer EM, Scovronick N, Armstrong B, Coelho MSZS, et al. Rapid increase in the risk of heat-related mortality. Nat Commun. 2023;14:4894.
10. Delforge D, Wathelet V, Below R, Lanfredi Sofia C, Tonnelier M, van Loenhout JAF, et al. EM-DAT: the Emergency Events Database. Int J Disaster Risk Reduct. 2025;124:105509.
11. Zhao Q, Li S, Ye T, Wu Y, Gasparrini A, Tong S, et al. Global, regional, and national burden of heatwave-related mortality from 1990 to 2019: a three-stage modelling study. PLoS Med. 2024;21:e1004364.
12. Centre for Research on the Epidemiology of Disasters (CRED). EM-DAT: The International Disaster Database - technical report. Brussels, Belgium: Université Catholique de Louvain; 2022.
13. Falchetta G, De Cian E, Sue Wing I, Carr D. Global projections of heat exposure of older adults. Nat Commun. 2024;15:3678.
14. Janoš T, Quijal-Zamorano M, Shartova N, Gallo E, Méndez Turrubiates RF, Beltrán Barrón ND, et al. Heat-related mortality in Europe during 2024 and health emergency forecasting to reduce preventable deaths. Nat Med. 2025;31:4065-74.
15. García-León D, Masselot P, Mistry MN, Gasparrini A, Motta C, Feyen L, et al. Temperature-related mortality burden and projected change in 1368 European regions: a modelling study. Lancet Public Health. 2024;9:e644-53.
16. Barriopedro D, Fischer EM, Luterbacher J, Trigo RM, García-Herrera R. The hot summer of 2010: redrawing the temperature record map of Europe. Science. 2011;332:220-4.
17. Sharma A, Andhikaputra G, Wang YC. Heatwaves in South Asia: characterization, consequences on human health, and adaptation strategies. Atmosphere. 2022;13:734.
18. World Weather Attribution. Climate change made the deadly heatwaves that hit millions of highly vulnerable people across Asia more frequent and extreme. World Weather Attribution; 2024.
19. Im ES, Pal JS, Eltahir EAB. Deadly heat waves projected in the densely populated agricultural regions of South Asia. Sci Adv. 2017;3:e1603322.
20. United Nations Economic and Social Commission for Asia and the Pacific (ESCAP). Asia-Pacific Disaster Report 2025: Rising Heat, Rising Risk. Bangkok, Thailand: ESCAP; 2025.
21. Fouillet A, Rey G, Wagner V, Laaidi K, Empereur-Bissonnet P, Le Tertre A, et al. Has the impact of heat waves on mortality changed in France since the European heat wave of summer 2003? A study of the 2006 heat wave. Int J Epidemiol. 2008;37:309-17.
22. Sheehan MC, Boned-Ombuena A, Cash-Gibson L, Damis-Wulff A, Fox MA. A global assessment of urban extreme weather early warning systems and public health engagement. Bull World Health Organ. 2025;103:294-303.
23. United Nations Office for Disaster Risk Reduction (UNDRR). Sendai Framework for Disaster Risk Reduction 2015-2030. Geneva, Switzerland: UNDRR; 2015.
24. Wu X, Wang J, Ge Y, Lai S, Zhang D, Ren Z, et al. Future heat-related mortality in Europe driven by compound day-night heatwaves and demographic shifts. Nat Commun. 2025;16:7420.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.70437
Eurasian J Emerg Med. 2026;25: 429-37

Fresh Frozen Plasma versus Prothrombin Complex Concentrate for Warfarin Reversal in the Emergency Department: A Retrospective Cohort Study of Clinical Outcomes and Costs

Halil İsa Çelik[1], Ertan Sönmez[2], Bahadır Taşlıdere[3], Suphi Bahadırlı[1], Yusuf Uğurlu[1], Mehmet Şam[1], Bedia Gülen[1]

[1]İstanbul Medipol University Faculty of Medicine, Department of Emergency Medicine, İstanbul, Türkiye
[2]Kütahya Health Sciences University Faculty of Medicine, Department of Emergency Medicine, Kütahya, Türkiye
[3]Bezmialem Vakif University Faculty of Medicine, Department of Emergency Medicine, İstanbul, Türkiye

Abstract

Aim: Rapid reversal of warfarin-associated coagulopathy is a frequent challenge in emergency departments (EDs). Four-factor prothrombin complex concentrates (PCCs) correct the international normalized ratio (INR) more rapidly than fresh frozen plasma (FFP), but their clinical and economic impact in routine practice, particularly in resource-limited settings, remains uncertain.

Materials and Methods: This retrospective cohort study was conducted in the ED of a tertiary university hospital. Adult warfarin-treated patients presenting with INR >1.2 and who received FFP alone or PCC (Cofact) alone for supratherapeutic INR and/or warfarin-associated bleeding were included. Clinical, laboratory, and direct product cost data were abstracted from electronic records. Primary outcomes were INR reduction (ΔINR) and achievement of target INR thresholds (≤1.5 and ≤1.8). Secondary outcomes included acute kidney injury (AKI), hospital and intensive care unit (ICU) length of stay, ED disposition, in-hospital mortality, and direct product costs. Multivariable logistic regression was used to explore independent associations.

Results: We analyzed 119 patients (92 FFP, 27 PCC). PCC was associated with a lower final INR and a greater ΔINR than FFP. PCC-treated patients were more likely to have gastrointestinal bleeding and markers of greater illness severity. AKI occurred more frequently in patients receiving PCC, and PCC use was independently associated with AKI after adjustment for age and presenting INR. PCC recipients had longer hospital stays, higher ICU utilization, and a numerically higher in-hospital mortality, although mortality estimates were imprecise. Both direct product costs and the cost per unit reduction in INR were substantially higher with PCC.

Conclusions: In this real-world ED cohort, PCC achieved superior laboratory reversal compared with FFP but was preferentially used in patients who were clinically sicker. While PCC was associated with higher AKI rates, longer hospital stays, and greater direct product costs, these findings likely reflect baseline illness severity and confounding by indication rather than a direct effect of the reversal agent. Our results are primarily hypothesis-generating and underscore the need for larger prospective, multicenter studies with comprehensive cost-effectiveness and safety analyses.

Keywords: Warfarin, plasma, fresh frozen, prothrombin complex concentrates, emergency service, hospital, acute kidney injury, costs and cost analysis

Corresponding Author: Prof. Bedia Gülen, İstanbul Medipol University Faculty of Medicine, Department of Emergency Medicine, İstanbul, Türkiye
E-mail: drbediagulen@yahoo.com **ORCID ID:** orcid.org/0000-0003-4307-2701

Cite this article as: Çelik Hİ, Sönmez E, Taşlıdere B, Bahadırlı S, Uğurlu Y, Şam M, et al. Fresh frozen plasma versus prothrombin complex concentrate for warfarin reversal in the emergency department: a retrospective cohort study of clinical outcomes and costs Eurasian J Emerg Med. 2026;25: 429-37

Received: 31.12.2025
Accepted: 06.09.2026
Published: 29.09.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Introduction

Warfarin is a long-standing oral anticoagulant widely used in the treatment of cardiovascular disease and in the prevention of thromboembolic events. Its efficacy in reducing morbidity and mortality has been demonstrated in conditions such as atrial fibrillation, mechanical heart valve prostheses, deep vein thrombosis, and pulmonary embolism (1-3). However, because warfarin has a narrow therapeutic window, requires careful dose adjustment, and is subject to numerous drug-drug and drug-food interactions, patients are at increased risk of both thromboembolism and bleeding (1,3,4).

Major bleeding is one of the most important clinical problems in patients presenting with warfarin-associated coagulopathy or supratherapeutic international normalized ratio (INR). Intracranial hemorrhage, gastrointestinal (GI) bleeding, and other bleeding foci that cause massive blood loss may result in death or permanent disability (5,6). Many warfarin-treated patients presenting to the emergency department (ED) are rapidly evaluated for bleeding or urgent invasive procedures, and fast and effective reversal of anticoagulation becomes a priority (5,7).

Fresh- frozen plasma (FFP) has been one of the most commonly used treatments for warfarin reversal. By providing coagulation factors, FFP aims to lower the INR, but it has important limitations, including large infusion volumes, prolonged administration time, volume overload, and transfusion-related reactions (5,6). In recent years, prothrombin complex concentrates (PCCs) have gained prominence as an alternative. Four-factor PCCs contain factors II, VII, IX, and X as well as natural anticoagulants, and have been shown to normalize INR more rapidly and predictably than FFP, with lower volumes and shorter administration times, which may be advantageous in hemodynamically fragile patients (7-10).

Randomized controlled trials and observational studies consistently demonstrate that PCCs are superior to FFP in terms of the speed and extent of INR correction. Some reports also suggest possible advantages in functional outcomes, although clear mortality benefits have not been consistently demonstrated, and concerns remain about thromboembolic complications and higher direct product costs (7,9,11). Given limited healthcare resources and the substantial cost burden of blood products, it is important to evaluate reversal strategies not only in terms of clinical efficacy but also with respect to cost and resource use, particularly in developing countries (5,9).

The aim of this study was to compare the clinical and economic outcomes of FFP versus PCC (Cofact) for anticoagulation reversal in warfarin-treated patients presenting to the ED with warfarin-associated coagulopathy. Specifically, we compared laboratory response (INR reduction), hospital and intensive care unit (ICU) lengths of stay, in-hospital mortality, development of acute kidney injury (AKI), and direct product costs between patients receiving FFP alone and those receiving PCC alone.

Materials and Methods

Study Design and Setting

We conducted a retrospective cohort study in the ED of a tertiary university hospital from 1 January 2015 to 31 December 2017. Using the hospital information system, we identified all adult patients receiving warfarin (Coumadin) whose presenting INR exceeded the laboratory upper reference limit of 1.2. A total of 1,130 warfarin-treated patients with INR >1.2 were identified; 821 did not receive specific reversal with FFP or PCC and were excluded.

Among the remaining 309 patients who received FFP and/or PCC for warfarin-related INR elevation, we included only those who received FFP alone or PCC alone and who had complete clinical, laboratory, and cost data, yielding a final cohort of 119 patients (92 FFP-only and 27 PCC-only). During the study period, the choice between FFP and PCC was not randomized but was left to the discretion of the treating physician; in practice, PCC tended to be used in more clinically severe cases or when rapid INR reversal was considered particularly important.

This study was approved by the Non-Interventional Clinical Research Ethics Committee of Bezmialem Vakıf University (approval date: 02.01.2018, decision/protocol no: 1/12). Due to the retrospective nature of the study, the requirement for informed consent was waived by the ethics committee.

Eligibility Criteria

Patients were eligible if they were aged ≥18 years, were current warfarin users, had a presenting INR>1.2, and received FFP alone or PCC alone in the ED for supratherapeutic INR and/or warfarin-associated bleeding. Warfarin-associated bleeding was defined as any overt bleeding event (e.g., GI, intracranial, genitourinary, or other sites) documented in ED records during warfarin therapy.

Patients were excluded if they were younger than 18 years, had no history of warfarin use, presented with an INR ≤1.2, had a bleeding diathesis without prior warfarin exposure, received both FFP and PCC or other hemostatic products in combination, or had missing key clinical, laboratory, or cost data.

Data Collection and Variables

Data were abstracted retrospectively from the hospital information system and laboratory databases using a standardized form.

For each patient, we recorded age, sex, presenting INR, post-treatment (last) INR, baseline and last serum creatinine levels, the presence of warfarin-associated bleeding and its anatomical location, treatment type (FFP vs PCC) and the units administered, ED disposition (discharge, ward admission, or ICU admission), in-hospital mortality, hospital length of stay, and direct product costs. Economic outcomes were limited to direct product acquisition costs (the price paid by the hospital pharmacy for FFP units and PCC vials). Indirect costs, such as personnel time, laboratory monitoring, and total hospital stay costs, were not included in this analysis.

INR measurements were performed using an Amax 200 Amelung analyzer (Trinity Biotech Plc., IDA Business Park, Bray, Co. Wicklow, Ireland). The "last INR" was defined as the INR value measured within the first 24 hours after initiation of FFP or PCC treatment, or earlier if the patient was discharged or died. Because the exact timing of follow-up INR measurements varied between patients, this value should be interpreted as a pragmatic, real-world indicator of early laboratory response rather than a strictly time-standardized endpoint. As a result, comparisons of the last INR and ΔINR between treatment groups may be influenced by variability in the timing of follow-up measurements. For each patient, we calculated the change in INR (ΔINR = presenting INR, last INR). While vitamin K administration and red blood cell transfusions were part of the standard clinical protocol for warfarin reversal and major bleeding management in our ED, these interventions were inconsistently documented in the electronic records, and therefore could not be included as independent covariates in the multivariable models.

Outcomes and Definitions

The primary laboratory outcome was the magnitude of INR reduction (ΔINR) and the proportion of patients achieving target INR thresholds (≤1.5 and ≤1.8). Clinical outcomes included the development of AKI, hospital and ICU lengths of stay, ED disposition, and in-hospital mortality. Economic outcomes were the total product cost per patient and the cost per unit of INR reduction.

AKI was defined as an increase in serum creatinine from baseline of ≥0.3 mg/dL or ≥1.5 times the baseline value during hospitalization, consistent with KDIGO criteria. Baseline creatinine was defined as the first measurement obtained in the ED. Due to the retrospective nature of the study, AKI was defined solely based on serum creatinine changes, as hourly urine output data were not consistently available in the electronic medical records to fulfill all KDIGO criteria. Hospital length of stay was calculated as the number of calendar days from ED presentation to discharge or death; same-day discharge was coded as 0 days.

Statistical Analysis

Continuous variables were assessed for normality and summarized as medians with interquartile ranges (IQRs). Categorical variables were summarized as counts and percentages. Comparisons between FFP and PCC groups were performed using the Mann-Whitney U test for continuous variables and the chi-square or Fisher's exact test for categorical variables.

We constructed multivariable logistic regression models to assess the independent association between treatment type (PCC vs FFP) and in-hospital mortality, AKI development, and ICU admission, adjusting for relevant covariates (age, presenting INR, presence of GI bleeding, and/or AKI, as appropriate). Given the limited number of events for mortality (n=5) and ICU admission (n=14), these models are underpowered and should be interpreted as exploratory. In particular, due to the very low number of in-hospital mortality events (n=5), the mortality model was interpreted with extreme caution, given the potential for overfitting and imprecise effect estimates. Two-sided p values <0.05 were considered statistically significant. No formal adjustment was made for multiple comparisons. Analyses were performed using IBM SPSS Statistics, version 25.0.

Results

Patient Characteristics

We included 119 patients who presented to the ED with warfarin-associated coagulopathy. A total of 92 patients received FFP and 27 received PCC (Cofact), each as the sole reversal agent. The groups were similar in age and sex distribution. The median age was 69.0 years (IQR 62.8-77.0) in the FFP group and 67.0 years (IQR 52.5-83.0) in the PCC group (p=0.87). Men constituted 46.7% (43/92) of the FFP group and 59.3% (16/27) of the PCC group (p=0.28) (Table 1).

Baseline serum creatinine values were comparable between groups, with medians of 1.10 mg/dL (IQR 1.10-1.40) in the FFP group and 1.10 mg/dL (IQR 1.05-1.45) in the PCC group (p=0.53). Last creatinine values were significantly higher in the PCC group than in the FFP group, with medians of 1.92 mg/dL (IQR 1.11-2.36) and 1.09 mg/dL (IQR 0.83-1.53), respectively (p=0.008). The change in creatinine was-0.08 (IQR -0.29-0.10) in the FFP group and 0.12 (IQR -0.15-0.99) in the PCC group (p=0.014). AKI occurred in 15.2% (14/92) of FFP-treated patients and 37.0% (10/27) of PCC-treated patients (p=0.026) (Table 1).

GI bleeding was common in both groups, but was more frequent in the PCC group, occurring in 70.7% (65/92) of FFP patients and 92.6% (25/27) of PCC patients (p=0.021).

Table 1. Baseline clinical and cost characteristics by treatment group

Characteristic	FFP (n=92)	PCC (n=27)	p value*
Demographics			
Age (years), median [IQR]	69.0 [62.8-77.0]	67.0 [52.5-83.0]	0.870
Gender			0.280
Male, n (%)	43 (46.7%)	16 (59.3%)	—
Female, n (%)	49 (53.3%)	11 (40.7%)	—
Renal function			
Baseline creatinine (mg/dL), median [IQR]	1.10 [1.10-1.40]	1.10 [1.05-1.45]	0.530
Last creatinine (mg/dL), median [IQR]	1.09 [0.83-1.53]	1.92 [1.11-2.36]	0.008
Creatinine change†, median [IQR]	-0.08 [-0.29-0.10]	0.12 [-0.15-0.99]	0.014
AKI present, n (%)‡	14 (15.2%)	10 (37.0%)	0.026
Coagulation parameters			
Presenting INR, median [IQR]	5.71 [3.52-7.77]	6.38 [4.09-13.39]	0.110
Last INR, median [IQR]	2.23 [1.76-2.91]	1.87 [1.52-2.01]	0.008
INR reduction (ΔINR)§, median [IQR]	3.14 [1.24-5.26]	3.91 [1.94-11.55]	0.039
Bleeding and disposition			
GI bleeding present, n (%)	65 (70.7%)	25 (92.6%)	0.021
Length of stay (days), median [IQR]	1.0 [0.0-6.0]	5.0 [2.5-8.0]	0.014
Discharged from ED, n (%)¶	44 (47.8%)	5 (18.5%)	0.023**
Ward admission, n (%)¶	39 (42.4%)	17 (63.0%)	
ICU admission, n (%)¶	9 (9.8%)	5 (18.5%)	
Mortality			
In-hospital mortality, n (%)	2 (2.2%)	3 (11.1%)	0.076
Product use and direct costs			
FFP units used, median [IQR]	2 [2-4]	—	—
FFP cost per patient (USD), median [IQR]	40 [40-80]	—	—
PCC units used, median [IQR]	—	2 [1-4]	—
PCC cost per patient (USD), median [IQR]	—	260 [130-520]	—
Total FFP cost (USD)	5,680	—	—
Total PCC cost (USD)	—	9,100	—

* Mann-Whitney U test was used for continuous variables, Fisher's exact test for binary categorical variables, and the chi-square test for the three-category disposition variable. † Continuous variable representing last creatinine minus baseline creatinine. ‡ Coded as 1 for patients who developed AKI and 0 for those who did not. §ΔINR = presenting INR - last INR. ¶Disposition variable: the p value corresponds to the chi-square test for the distribution across the three categories (discharge, ward, ICU). ** Significant difference in the overall distribution of disposition. AKI: Acute kidney injury, ED: Emergency department, FFP: Fresh frozen plasma, GI: Gastrointestinal, ICU: Intensive care unit, INR: International normalized ratio, IQR: Interquartile range, PCC: Prothrombin complex concentrate, USD: United States dollars

Laboratory Response and Clinical Outcomes

INR values tended to be higher in the PCC group, with medians of 5.71 (IQR 3.52-7.77) in the FFP group and 6.38 (IQR 4.09-13.39) in the PCC group; however, this difference was not statistically significant (p=0.11). Final post-treatment INR values were significantly lower in the PCC group [median 1.87 (IQR 1.52-2.01)] compared with the FFP group [median 2.23 (IQR 1.76-2.91)] (p=0.008). The reduction in INR (ΔINR) was greater among PCC-treated patients [median 3.91 (IQR 1.94-11.55)] than among FFP-treated patients [median 3.14 (IQR 1.24-5.26)] (p=0.039).

More PCC-treated patients achieved target INR thresholds, although the differences were not statistically significant. INR ≤1.5 was reached in 7.6% (7/92) of FFP patients and 22.2% (6/27) of PCC patients (p=0.071), and INR ≤1.8 was reached in 26.1% (24/92) and 40.7% (11/27), respectively (p=0.156) (Tables 1 and 2).

Median total hospital length of stay was significantly longer in the PCC group (5.0 days; IQR 2.5-8.0) than in the FFP group (1.0 day; IQR 0.0-6.0) (p=0.014). Among patients with GI bleeding, the median length of stay was 0 days (IQR 0-4) in the FFP group and 5 days (IQR 3-8) in the PCC group (p<0.001), whereas among

those without GI bleeding, the length of stay did not differ significantly between groups. Across the entire cohort, patients who developed AKI had longer hospital stays than those without AKI (5.5 days [IQR 1-9] vs 3.0 days [IQR 0-6]; p = 0.026) (Tables 1 and 2).

Disposition from the ED differed between treatment groups. Among FFP-treated patients, 47.8% (44/92) were discharged directly from the ED, 42.4% (39/92) were admitted to a general ward, and 9.8% (9/92) were admitted to the ICU. Among PCC-treated patients, 18.5% (5/27) were discharged, 63.0% (17/27) were admitted to a ward, and 18.5% (5/27) were admitted to the ICU. The overall distribution of disposition categories differed significantly between groups (p=0.023), with higher rates of ward and ICU admission in the PCC group. ICU admission rates were numerically higher in the PCC group, but did not differ significantly (9.8% vs 18.5%; p=0.305). ICU admission did not differ significantly between patients with and without GI bleeding; however, patients who developed AKI were more likely to require ICU care (25.0% vs 8.4%; p=0.035) (Tables 1 and 2).

In-hospital mortality occurred in 2.2% (2/92) of FFP-treated patients and 11.1% (3/27) of PCC-treated patients (p=0.076). In a multivariable logistic regression model including treatment type, age, and presenting INR, PCC treatment was associated with higher odds of mortality (odds ratio (OR) 5.82; 95% confidence interval (CI) 0.86-39.30; p=0.071]. However, this association did not reach statistical significance and had a very wide confidence interval, due to the low number of events (n=5). Age and presenting INR were not independently associated with mortality (Tables 1 and 3).

In the multivariable model for AKI, after adjustment for age and presenting INR, PCC use remained significantly associated with AKI compared with FFP (OR 5.21; 95% CI 1.71-15.86; p=0.004).

Table 2. Subgroup analyses of INR targets, length of stay, and ICU admission

Analysis/subgroup	FFP	PCC (Cofact)	p value*
INR targets			
Reached INR ≤1.5, n/N (%)	7/92 (7.6%)	6/27 (22.2%)	0.071
Reached INR ≤1.8, n/N (%)	24/92 (26.1%)	11/27 (40.7%)	0.156
Length of stay (days), median [IQR]			
GI bleeding present	0 [0-4] (n=65)	5 [3-8] (n=25)	**<0.001**
GI bleeding absent	5 [0-9] (n=27)	4 [2-6] (n=2)	0.759
AKI present	5.5 [1.5-9.0] (n=14)	5.5 [1.25-7.5] (n=10)	0.840
AKI absent	0.0 [0.0-5.0] (n=78)	4.0 [3.0-8.0] (n=17)	**<0.001**
ICU admission, n/N (%)			
By treatment			
FFP	9/92 (9.8%)	—	0.305
PCC	—	5/27 (18.5%)	
By GI bleeding status (entire cohort)‡			
GI bleeding present	10/90 (11.1%)		0.743
GI bleeding absent	4/29 (13.8%)		
By AKI status (entire cohort)‡			
AKI present	6/24 (25.0%)		0.035
AKI absent	8/95 (8.4%)		

*Mann-Whitney U test was used for continuous variables; Fisher's exact test was used for proportions. ‡For the entire cohort (FFP + PCC combined), comparison of ICU admission rates by GI bleeding status (present vs absent) and by AKI status (present vs absent). AKI: Acute kidney injury, FFP: Fresh frozen plasma, GI: Gastrointestinal, ICU: Intensive care unit, INR: International normalized ratio, IQR: Interquartile range, PCC: Prothrombin complex concentrate

Table 3. Multivariable predictors of in-hospital mortality

Variable	OR	95% CI (Lower-Upper)	p value
Treatment: PCC (Cofact) vs FFP*	5.82	0.86-39.30	0.071
Age (per 1-year increase)	0.98	0.93-1.04	0.554
Presenting INR (per 1-unit increase)	0.95	0.76-1.18	0.632

Dependent variable: in-hospital mortality (0 =alive, 1 =death). Model: mortality ~ treatment_cof + age + presenting INR. Model n=119; number of events =5. CI: Confidence interval, FFP: Fresh frozen plasma, INR: International normalized ratio, OR: Odds ratio, PCC: Prothrombin complex concentrate. *Reference category: FFP

However, this association should be interpreted with extreme caution due to the significant potential for residual confounding and indication bias inherent in this non-randomized, retrospective cohort. Specifically, the model could not account for several critical determinants of renal risk such as baseline comorbidities, hemodynamic instability, clinical bleeding severity, total transfusion requirements, or potential nephrotoxic exposures which may have been more prevalent in the PCC group. Regarding other covariates, increasing age was associated with a modest but significant increase in AKI risk (OR 1.04 per year; 95% CI 1.00-1.08; p=0.029), whereas the presenting INR showed only a borderline association (p=0.067) (Table 4). In a separate model evaluating predictors of ICU admission, neither treatment type nor GI bleeding reached statistical significance; however, AKI demonstrated a trend toward significance as an independent predictor of ICU admission (OR 3.25; 95% CI 0.96-11.00; p=0.058) (Table 5).

Cost Outcomes

In the cost analysis restricted to product costs, median total cost per patient was 40 United States dollars (USD) (IQR 40-80; mean 61.7±42.3) in the FFP group and 260 USD (IQR 130-520; mean 337.0±245.8) in the PCC group (p<0.001). Total expenditure on FFP was 5,680 USD, whereas total expenditure on PCC was 9,100 USD. The cost per unit of INR reduction was also higher in the PCC group, with a median of 69.1 USD per INR unit (IQR 30.4-157.7; mean 104.3±98.7), compared with 16.0 USD per unit (IQR 8.9-51.1; mean 45.1±68.2) in the FFP group (p<0.001). When stratified by AKI, the median total product cost was 100 USD (IQR 40-130) in patients who developed AKI and 60 USD (IQR 40-120) in those who did not, representing a difference of borderline statistical significance (p=0.082) (Table 6).

Table 4. Multivariable predictors of acute kidney injury

Variable	OR	95% CI (Lower-Upper)	p value
Treatment: PCC (Cofact) vs FFP*	5.21	1.71-15.86	0.004
Age (per 1-year increase)	1.04	1.00-1.08	0.029
Presenting INR (per 1-unit increase)	0.88	0.77-1.01	0.067

Model details: Dependent variable: AKI (0 =no, 1 =yes). Model: AKI ~ treatment_cof + age + presenting INR. Model n=119; number of events = 24. AKI: Acute kidney injury, CI: Confidence interval, FFP: Fresh frozen plasma, INR: International normalized ratio, OR: Odds ratio, PCC: Prothrombin complex concentrate. *Reference category: FFP

Table 5. Multivariable predictors of ICU admission

Variable	OR	95% CI (Lower-Upper)	p value
Treatment: PCC (Cofact) vs FFP*	1.77	0.48-6.56	0.390
AKI present vs absent†	3.25	0.96-11.00	0.058
GI bleeding present vs absent‡	0.63	0.17-2.38	0.493

Model details: Dependent variable: ICU admission (0 =no, 1 =yes). Model: ICU admission ~ treatment_cof + AKI + GI bleeding. Model n=119; number of events =14. AKI: Acute kidney injury, CI: Confidence interval, FFP: Fresh frozen plasma, GI: Gastrointestinal, ICU: Intensive care unit, OR: Odds ratio, PCC: Prothrombin complex concentrate. *Reference category: FFP, †Coded as 1 for AKI present and 0 for AKI absent, ‡Coded as 1 for GI bleeding present and 0 for GI bleeding absent

Table 6. Cost outcomes by treatment and AKI status

Analysis/group	Median [IQR]	Mean ± SD	p value*
Total cost per patient (USD)			
FFP (n=92)	40 [40-80]	61.7±42.3	
PCC (n=27)	260 [130-520]	337.0±245.8	<0.001
Cost per unit INR reduction (USD per INR unit)†			
FFP (n=92)	16.0 [8.9-51.1]	45.1±68.2	
PCC (n=27)	69.1 [30.4-157.7]	104.3±98.7	<0.001
Total cost by AKI status (entire cohort)			
AKI present (n=24)	100 [40-130]	95.0±78.4	0.082
AKI absent (n=95)	60 [40-120]	112.6±156.3	

*Mann-Whitney U test. Total cost = cost of the respective reversal product (FFP or PCC), †Cost per unit INR reduction = total product cost / ΔINR (patients with ΔINR > 0 included). AKI: Acute kidney injury, FFP: Fresh frozen plasma, INR: International normalized ratio, IQR: Interquartile range, PCC: Prothrombin complex concentrate, SD, standard deviation, USD, United States dollars

Discussion

In this retrospective cohort of warfarin-treated patients requiring reversal in a tertiary ED, PCC (Cofact®) achieved greater INR reduction and lower post-treatment INR than FFP, consistent with prior randomized and observational studies (7-11). However, PCC use in our real-world cohort was associated with higher rates of AKI, longer hospital stays, higher direct product costs, and numerically higher, but statistically inconclusive, mortality compared with FFP.

Our findings align with established evidence demonstrating that four-factor PCC provides faster and more complete correction of warfarin-induced coagulopathy than FFP. In the phase IIIb plasma-controlled trial by Sarode et al. (8), most PCC-treated patients achieved target INR within 30-60 minutes, whereas INR normalization was slower and less complete with plasma. Goldstein et al. (10) similarly reported that PCC provided more rapid INR normalization and comparable safety in patients requiring urgent surgical or invasive interventions. In our cohort, PCC recipients had a larger median ΔINR and lower final INR, reinforcing the laboratory advantage of PCC observed in the literature (7-10).

Despite superior INR correction, PCC has not consistently translated into clear survival or functional benefits over FFP. Randomized trials generally showed comparable mortality and thromboembolic event rates between PCC and plasma, with the main benefit of PCC being faster and more reliable reversal (8-11). Meta-analytic data also suggest that PCC is safe and effective for vitamin K antagonist reversal, with no substantial absolute increase in thromboembolic complications, though careful patient selection is emphasized, especially in high-risk individuals (7,11). In our study, the small number of deaths limited statistical power and yielded wide confidence intervals for the mortality estimate. PCC-treated patients had a numerically higher mortality, but this association did not reach conventional statistical significance and may reflect confounding by indication, with PCC preferentially used in more severely ill or high-risk patients.

One of the most notable findings in our cohort was a higher incidence of AKI among PCC-treated patients; this association remained statistically significant even after adjustment for age and presenting INR. However, this finding must be interpreted with caution due to substantial confounding by indication. In our clinical practice, treatment allocation was not randomized, and PCC was preferentially selected for patients perceived to be more severely ill. This is evidenced by the significantly higher prevalence of GI bleeding in the PCC group (92.6% vs 70.7%, p=0.021) and by a trend toward higher presenting INR values (median 6.38 vs 5.71).

Rather than indicating a direct nephrotoxic effect of the product, these baseline differences specifically the greater hemodynamic instability and clinical complexity associated with extensive bleeding are primary drivers of both pre-renal and intrinsic kidney injury. Since the retrospective nature of our study precluded adjustment for all markers of physiological stress and transfusion requirements, the observed incidence of AKI likely reflects illness severity in patients requiring urgent reversal with PCC. We emphasize that this indication bias is an inherent limitation of the non-randomized, retrospective design and cannot be fully eliminated by statistical adjustment. Only a randomized comparison could determine whether the observed differences between PCC and FFP reflect a true treatment effect rather than being due to the baseline severity of patients selected to receive PCC.

Warfarin-associated bleeding and shock states, as well as underlying comorbidities and concurrent nephrotoxic medications, are recognized contributors to AKI in this population. The literature does not generally implicate PCC as intrinsically nephrotoxic; instead, AKI is more often related to the severity of illness and hemodynamic instability (7). The strong association between AKI and both longer hospital stays and higher ICU admission rates further supports the interpretation of AKI as a marker of overall illness severity rather than a consequence of the treatment choice itself.

Our data also showed substantial differences in direct product costs between FFP and PCC. Economic considerations are increasingly important in decisions about warfarin reversal strategies. PCC is substantially more expensive per unit than FFP, yet may reduce overall costs by shortening the length of stay and reducing transfusion-related resource use, particularly in high-risk subgroups, such as patients with intracranial hemorrhage. Khorsand et al. (9), using a Dutch health system perspective, suggested that the higher acquisition costs of PCC might potentially be offset by reductions in transfusion requirements and hospitalization-related resource utilization when broader healthcare costs are considered. In our study, the economic analysis was restricted to direct product acquisition costs-the price paid by the hospital pharmacy for FFP units and PCC vials. From this limited perspective, both the total cost per patient and the cost per unit reduction in INR were markedly higher in the PCC group. However, it must be clarified that this represents a restricted cost comparison rather than a comprehensive cost-effectiveness analysis. By excluding broader healthcare expenses such as personnel time, laboratory monitoring, and specifically

the costs associated with ward and ICU stays, we could not fully capture the potential cost-offset profile of PCC (7,9,11,12). Given that PCC recipients in our cohort experienced significantly longer hospitalizations, any potential economic advantage from faster INR correction may have been outweighed by the costs of managing patients who were more clinically complex and more severely ill. These findings underscore the necessity for future studies to incorporate total hospital costs and complication-related expenses to provide a more definitive economic evaluation. Because PCC recipients in our cohort had longer hospital and ICU stays, the true difference in overall care costs between PCC and FFP cannot be determined from acquisition costs alone, and our cost findings should therefore not be interpreted as evidence that PCC is less cost-effective than FFP overall.

Clinically, FFP is disadvantaged by the need for larger infusion volumes, longer administration times, and the risk of transfusion-associated circulatory overload and acute lung injury (5,7). PCC, in contrast, can be administered rapidly in much smaller volumes, offering pragmatic advantages for unstable ED patients and those with limited cardiac reserve. Our data do not include systematic documentation of transfusion reactions, volume overload, or thromboembolic events, limiting our ability to compare safety profiles comprehensively. As such, our conclusions regarding safety are based mainly on AKI, length of stay, and ICU utilization, all of which are heavily influenced by baseline severity of illness rather than by the reversal product alone (7,11,12).

Study Limitations

This study has several limitations that warrant consideration. First, its retrospective single-center design and modest sample size particularly in the PCC group limit both internal validity and generalizability. Second, treatment allocation was not randomized, leading to substantial confounding by indication, as PCC was preferentially used in clinically more severe cases. Because PCC recipients differed systematically from FFP recipients in baseline illness severity, our comparative findings on AKI, length of stay, and costs may not be generalizable to settings in which treatment selection is less strongly linked to clinical severity. Furthermore, detailed information on baseline comorbidities and precise bleeding severity scores was not systematically available. Notably, the inconsistent documentation of vitamin K administration and red blood cell transfusions represents a critical limitation. Since these interventions are standard components of warfarin reversal and major bleeding management, their omission from the multivariable models may have influenced the observed rates of INR correction and clinical recovery. Finally, the economic analysis was restricted to direct product acquisition costs and did not account for the broader financial impact of ICU stays or of the management of complications.

We also lacked systematic data on thromboembolic events, transfusion reactions, and volume overload; therefore, the safety comparison between FFP and PCC is necessarily incomplete. AKI assessment was limited to serum creatinine-based criteria because urine-output data and complete KDIGO staging parameters were not consistently available in the retrospective electronic medical records. The economic analysis was restricted to direct product acquisition costs of FFP and PCC excluded ward or ICU stays, procedures, imaging, or readmissions. Multiple outcomes and subgroup analyses were performed without adjustment for multiple comparisons; therefore, secondary findings should be interpreted as exploratory and hypothesis-generating.

Conclusion

Among warfarin-treated patients presenting to the ED with supratherapeutic INR and/or warfarin-associated bleeding, 4-factor PCC (Cofact®) achieved more rapid and pronounced INR correction than FFP, consistent with previous randomized and observational studies. In this real-world ED cohort, 4-factor PCC achieved superior laboratory reversal compared with FFP but was administered to clinically sicker patients and was associated with higher AKI rates, longer hospitalizations, and greater direct product costs. Given the small number of in-hospital mortality events, our findings were underpowered to detect definitive survival differences and should therefore be interpreted with caution. The observed association between PCC treatment and AKI is likely driven by confounding by indication and baseline illness severity rather than a direct nephrotoxic effect of the reversal agent. These findings are primarily hypothesis-generating and underscore the need for larger prospective, multicenter studies with comprehensive cost-effectiveness and safety analyses to clarify the clinical impact of PCC across different warfarin-related bleeding scenarios.

Given the retrospective, non-randomized design, modest sample size particularly in the PCC group and incomplete data on key clinical confounders and safety endpoints, these findings should be interpreted with caution and regarded primarily as hypothesis-generating. Decisions to use PCC versus FFP in the ED should balance the need for rapid INR normalization with the patient's baseline clinical status, renal risk, bleeding site, and local economic constraints. Larger prospective multicenter studies with standardized treatment protocols and comprehensive cost-effectiveness and safety analyses are needed to clarify the clinical and economic impact of PCC in different warfarin-related bleeding scenarios.

Ethics

Ethics Committee Approval: This study was approved by the Non-Interventional Clinical Research Ethics Committee of Bezmialem Vakıf University (approval date: 02.01.2018, decision/protocol no: 1/12).

Informed Consent: Due to the retrospective nature of the study, the requirement for informed consent was waived by the ethics committee.

Author Contributions

Concept: H.İ.Ç., E.S. Design: E.S., B.G., Data Collection or Processing: H.İ.Ç., Y.U., Analysis or Interpretation: H.İ.Ç., S.B., E.S., Literature Search: H.İ.Ç., M.Ş., B.G., B.T., Writing: H.İ.Ç., B.T., B.G.,

Footnotes

Conflict of Interest: Bahadır Taşlıdere is an Associate Editor of the Eurasian Journal of Emergency Medicine but had no role in the editorial process. Independent editors evaluated the manuscript. The other authors declared no conflicts of interest.

Financial Disclosure: The author declared that this study received no financial support.

References

1. Hirsh J, Fuster V, Ansell J, Halperin JL. American Heart Association/American College of Cardiology Foundation guide to warfarin therapy. Circulation. 2003;107:1692-711.
2. January CT, Wann LS, Calkins H, Chen LY, Cigarroa JE, Cleveland JC Jr, et al. 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation. Circulation. 2019;140:e125-e51.
3. Holbrook A, Schulman S, Witt DM, Vandvik PO, Fish J, Kovacs MJ, et al. Evidence-based management of anticoagulant therapy. Chest. 2012;141(2 Suppl):e152S-e84S.
4. Pollack CV Jr, Reilly PA, Eikelboom J, Glund S, Verhamme P, Bernstein RA, et al. Idarucizumab for dabigatran reversal—full cohort analysis. N Engl J Med. 2017;377:431-41.
5. Makris M, Van Veen JJ, Maclean R, Watson HG. Warfarin anticoagulation reversal: management of the asymptomatic and bleeding patient. J Thromb Thrombolysis. 2010;29:171-81.
6. Holland L, Warkentin TE, Refaai MA, Crowther MA, Johnston MA, Sarode R. Suboptimal effect of a three-factor prothrombin complex concentrate (Profilnine-SD) in correcting supratherapeutic international normalized ratio due to warfarin overdose. Transfusion. 2009;49:1171-7.
7. Dentali F, Marchesi C, Giorgi Pierfranceschi M, Crowther M, Garcia D, Hylek E, et al. Safety of prothrombin complex concentrates for rapid anticoagulation reversal of vitamin K antagonists: a meta-analysis. Thromb Haemost. 2011;106:429-38.
8. Sarode R, Milling TJ Jr, Refaai MA, Mangione A, Schneider A, Durn BL, et al. Efficacy and safety of a 4-factor prothrombin complex concentrate in patients on vitamin K antagonists presenting with major bleeding: a randomized, plasma-controlled, phase IIIb study. Circulation. 2013;128:1234-43.
9. Khorsand N, Veeger NJGM, van Hest RM, Muller M, Meijer K. Prothrombin complex concentrate (PCC) versus fresh frozen plasma (FFP) for warfarin reversal: a cost analysis. Thromb Res. 2014;133:2-7.
10. Goldstein JN, Refaai MA, Milling TJ Jr, Lewis B, Goldberg-Alberts R, Hug BA, Sarode R. Four-factor prothrombin complex concentrate versus plasma for rapid vitamin K antagonist reversal in patients needing urgent surgical or invasive interventions: a phase 3b, open-label, non-inferiority, randomised trial. Lancet. 2015;385:2077-87.
11. Milling TJ Jr, Sarode R. Reversal of warfarin with four-factor prothrombin complex concentrate in patients with major bleeding: A review of hemostatic efficacy and safety. Thromb Haemost. 2015;113:931-42.
12. Freeman WD, Brott TG, Barrett KM, Castillo PR, Deen HG Jr, Zubkov AY, et al. Recombinant activated factor VII for rapid reversal of warfarin anticoagulation in acute intracranial hemorrhage. Mayo Clin Proc. 2004;79:1495-500.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.61224
Eurasian J Emerg Med. 2026;25: 438-48

Evaluating the Quality of Information in YouTube Videos on Alcohol Intoxication and Poisoning: Implications for Emergency Medicine

Hülya Yılmaz Başer[1], Sema Ayten[2]

[1]Bandırma Onyedi Eylül University Faculty of Medicine, Department of Emergency Medicine, Balıkesir, Türkiye
[2]İstanbul Medeniyet University Göztepe Training and Research Hospital, Prof. Dr. Süleyman Yalçın City Hospital, Department of Emergency Medicine, İstanbul, Türkiye

Abstract

Aim: This study evaluated, from an emergency medicine perspective, the quality, reliability, and first-intervention content of YouTube videos on alcohol-related intoxication and poisoning

Materials and Methods: On October 1, 2023, YouTube was searched for "alcohol intoxication" and "alcohol poisoning." The first 50 eligible English-language videos were included in this cross-sectional study. Two experienced emergency medicine physicians jointly evaluated the videos and recorded consensus scores using DISCERN, the Global Quality score (GQS), and the Journal of the American Medical Association (JAMA) benchmark criteria. Non-parametric tests were used for comparisons and correlations.

Results: The videos had accumulated 21,254,061 views. Symptoms were described in 30 videos (60%), whereas first-intervention information was provided in 19 videos (38%). Rating data were available for 49 videos; the median (interquartile range) Dissemination of Information on Consumer Health (DISCERN), GQS, and JAMA scores were 27 (8), 3 (2), and 2 (1), respectively. Scores did not differ significantly among academic, personal/non-academic, and commercial sources ($p=0.207$, $p=0.271$, and $p=0.923$, respectively). Videos describing symptoms had higher DISCERN scores (adjusted $p=0.002$), whereas those providing first-intervention information had higher DISCERN and GQS scores (adjusted $p<0.001$ and $p=0.003$, respectively). DISCERN and GQS were strongly correlated (rho=0.820, Holm-adjusted $p<0.001$).

Conclusion: YouTube videos on alcohol-related intoxication and poisoning frequently lack actionable emergency information. No statistically significant difference in quality scores was detected across uploader sources; however, the small and unequal group sizes preclude drawing conclusions of equivalence. Symptom content and first-intervention content were associated with higher quality scores. Public-facing videos should provide clear and timely guidance for seeking emergency medical assistance.

Keywords: Alcoholic intoxication, poisoning, emergency service, social media, health communication, YouTube

Introduction

Despite the potential benefits of platforms like YouTube, previous studies have shown that substance-related videos, including those promoting cannabis and synthetic cannabinoids, often depict recreational use and are frequently shared with minimal age restrictions (1). This increases the risk of exposure to health misinformation, particularly among younger and more vulnerable audiences. Similarly, alcohol consumption appears to be rising in today's society, influenced in part by the widespread reach and impact of social media (2). Alcohol is the most widely used intoxicating substance worldwide, and the consequences of unhealthy alcohol use are correspondingly substantial (3). Alcohol consumption exhibits a broad distribution across the general population and has a significant impact on healthcare systems (4). Alcohol consumption, reported to be involved in 5.9%

Corresponding Author: Ass. Prof. Hülya Yılmaz Başer, Bandırma Onyedi Eylül University Faculty of Medicine, Department of Emergency Medicine, Balıkesir, Türkiye
E-mail: ylmz_hly_35@yahoo.com **ORCID ID:** orcid.org/0000-0002-1416-1521
Cite this article as: Yılmaz Başer H, Ayten S. Evaluating the quality of information in YouTube videos on alcohol intoxication and poisoning: implications for emergency medicine. Eurasian J Emerg Med. 2026;25: 438-48

Received: 04.07.2026
Accepted: 08.09.2026
Published: 29.09.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

of all emergency department (ED) visits in Ireland, represents a significant global public health concern, contributing to one in five ED presentations and one in six hospital admissions worldwide (5,6). The 61.6% increase in alcohol-related ED visits and the 47% rise in visits specifically due to alcohol intoxication in the United States between 2006 and 2014 underscore the growing burden of alcohol misuse as a public health issue (7).

Alcohol-related poisoning may result from excessive ethanol intake or the ingestion of toxic alcohols, including methanol and ethylene glycol, which may be present in illicit alcoholic beverages, cleaning products, and automotive fluids (8,9). Both methanol and ethylene glycol have caused numerous mass-poisoning outbreaks (10). Clinical features used in the diagnosis of acute ethanol intoxication include a history of ethanol consumption, behavioral changes, slurred speech, impaired coordination, nystagmus, memory impairment, and the absence of another condition that adequately explains the clinical presentation (11,12). In recent years, the increased accessibility of medical information on the internet has led individuals to use web resources not only for information on the perceived positive aspects of alcohol use but also for information regarding intoxication-related situations, which indicates possible reliance on social media. Among these platforms, YouTube, owned by Google since 2006, is one of the most widely used video-sharing platforms (13). Given the prevalence of alcohol-related emergencies and the growing reliance on social media platforms for medical information, evaluating the quality of social media content has become a public health priority.

This study aimed to characterize the content, quality, reliability, and information on first interventions in YouTube videos addressing alcohol-related intoxication and poisoning, using internationally accepted assessment instruments.

Materials and Methods

On October 1, 2023, YouTube was searched using the terms "alcohol intoxication" and "alcohol poisoning." The search was conducted from Türkiye using an incognito browser session to minimize the influence of previous search and viewing history. No sorting, upload-date, language, or geographic filters were applied during the search, and the results were displayed according to YouTube's default relevance-based ranking. The search results generated using the two terms were screened sequentially, and the first 50 unique eligible videos were included. Videos were considered eligible if they addressed alcohol intoxication or poisoning and were available in English through either English-language audio or English subtitles. The search results and relevant video characteristics were recorded on the same day to limit the effect of subsequent changes in YouTube rankings and engagement metrics.

The terms "alcohol intoxication" and "alcohol poisoning" were used as complementary search terms to maximize the retrieval of relevant public-facing videos rather than as interchangeable clinical diagnoses. The primary focus of the study was acute ethanol intoxication; however, videos addressing toxic alcohol poisoning, such as methanol or ethylene glycol, were also retained when they contained information relevant to emergency recognition or intervention. During content review, videos specifically addressing methanol, ethylene glycol, or other toxic alcohols were also identified and counted. These videos were not classified as a separate analytical category. The binary "first-intervention information" variable was applied to all included videos to indicate whether each video presented any immediate emergency action or medical management relevant to the type of intoxication discussed.

This study is an observational, cross-sectional analysis of publicly available YouTube videos related to alcohol intoxication and poisoning. No interventions, experiments, or participant interactions were conducted. Data were collected passively, without altering the video content or influencing its ranking on the platform. Changes in video content, ranking, or engagement metrics over time were not evaluated.

The names of individuals and organizations creating the videos were kept confidential. Since all videos were publicly available on the social media website YouTube.com and no human or animal participants were involved, ethical committee approval was not required.

Video Characteristics, Content, Quality, and Reliability Assessment

The videos were characterized by duration (minutes), views, days since upload to YouTube, comments, and likes. Categorical characteristics included: video source (academic association/group, personal/non-academic, or commercial/advertisement), country of upload, target audience (physicians or non-physicians/general public), and language format (English audio or English subtitles). Videos were also assessed as to whether they introduced types of alcohol, described the signs or symptoms of alcohol-related intoxication or poisoning, and provided information on initial intervention.

Information on first intervention was recorded as a binary variable (present/absent). A video was classified as providing first-intervention information if it included at least one actionable recommendation concerning the immediate response to or early medical management of suspected alcohol-related intoxication or poisoning. Relevant recommendations included activation

of emergency medical services or urgent referral to an ED assessment or support of airway, breathing, and circulation; protective positioning and measures to reduce aspiration risk; observation and monitoring; avoidance of potentially unsafe responses, such as inducing vomiting; or other scenario-appropriate early medical management. For videos addressing toxic alcohols, urgent medical evaluation and information on appropriate antidotal or enhanced-elimination treatments were also considered relevant as first-intervention content. Videos that only described symptoms, long-term consequences, prevention, or general information and lacked an actionable recommendation for acute management were classified as not providing first-intervention information. These criteria were informed by established emergency medicine recommendations for the management of acute alcohol intoxication and substance intoxication (3,12). This binary variable reflected the presence of relevant first-intervention information rather than the completeness of the intervention, the full guideline concordance, or the clinical accuracy of every recommendation provided.

Like rate and view rate were calculated to standardize engagement metrics across videos. The Like rate was defined as the number of likes per 100 video views and calculated as [(number of likes/number of views) × 100]. The view rate was expressed as the annualized number of video views and calculated as [(number of views × 365)/number of days since upload].

Standardized scoring instruments were used to evaluate the quality and reliability of the videos. The Journal of the American Medical Association (JAMA) benchmark criteria, developed by Silberg et al. (14), were used to assess the transparency and reliability of online medical information. The JAMA criteria evaluate four domains: authorship, attribution or citation, disclosure, and currency, producing a total score ranging from 0 to 4. The Global Quality score (GQS), developed by Singh et al. (15), was used to assess the overall quality, educational value, flow, and usefulness of each video on a 5-point Likert scale. JAMA and GQS provide general, topic-independent assessments of online health information (16). The Dissemination of Information on Consumer Health (DISCERN) instrument, developed by Charnock et al. (17), was used to assess the reliability and quality of information concerning treatment choices. The 15-item DISCERN instrument used in the present study rated each item on a 1-5 scale, producing a total score ranging from 15 to 75; higher scores indicating better-quality information.

All eligible videos were viewed and evaluated simultaneously by two emergency medicine physicians, each with approximately 10 years of experience in emergency medicine. One evaluator was a faculty member, and the other was a practicing emergency medicine specialist. Before the assessment, the two evaluators reviewed the criteria of the DISCERN, GQS, and JAMA instruments to establish a common interpretive framework. Each video was viewed in full and scored jointly by the two evaluators. Following each assessment, a joint consensus score was entered for each instrument. Because the ratings were generated by joint consensus rather than independently, inter-rater reliability coefficients could not be calculated. During final data verification, the archived analytical dataset was found to be missing the complete consensus rating record for one video in the personal/non-academic source category. The missing data involved all three instruments (DISCERN, GQS, and JAMA). When re-evaluation was attempted during the revision process, the video was no longer accessible on YouTube; consequently, its scores could not be reconstructed reliably. The video was retained in descriptive and content-based analyses using information recorded on the original search date, but was excluded from all score-based analyses. No missing rating scores were imputed.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics version 22.0 (IBM Corp., Armonk, NY, USA). The revised analyses, including post hoc, exact, and multiple-comparison procedures, were independently verified using Python 3 and SciPy version 1.17.0. Continuous and ordinal variables were summarized using the median and interquartile range (IQR), with the IQR defined as the difference between the 75th and 25th percentiles. Categorical variables were presented as counts and percentages. The distribution of quantitative variables was assessed using the Shapiro-Wilk test. Because the relevant variables were not normally distributed and some scoring variables were ordinal, non-parametric statistical methods were used.

Comparisons of quantitative or ordinal variables across the three video-source categories were performed using the Kruskal-Wallis test. When the overall Kruskal-Wallis test was statistically significant, Dunn's post hoc test with Bonferroni adjustment was used for pairwise comparisons. Because several comparisons involved small expected cell counts, categorical variables were compared across video-source categories using the Fisher-Freeman-Halton exact test. When an overall Fisher-Freeman-Halton test was statistically significant, pairwise comparisons were performed using two-sided Fisher's exact tests with Holm adjustment.

Spearman's rank correlation coefficient was used to assess associations among quantitative and ordinal variables. Non-ordinal nominal variables (including video source) and binary categorical variables indicating the presence or absence of specific content were not included in the correlation analysis. Correlation strength was interpreted according to the absolute value of the

coefficient as negligible (0.00-0.10), weak (0.10-0.39), moderate (0.40-0.69), strong (0.70-0.89), or very strong (0.90-1.00), based on the classification proposed by Schober et al. (18). To account for multiple testing in Table 3, the Holm procedure was applied across the 15 unique correlation tests. Holm-adjusted two-sided p values are reported; an adjusted p value <0.05 is considered statistically significant.

To evaluate whether video quality differed based on the presence of specific content, DISCERN, GQS, and JAMA scores were compared using the Mann-Whitney U test between videos with and without information on alcohol types, intoxication or poisoning symptoms, and first intervention. Because nine comparisons were performed for these content characteristics, the resulting p values were adjusted using the Holm procedure. Missing data were handled using an available-case approach, and no values were imputed. Accordingly, analyses involving DISCERN, GQS, or JAMA scores were based on the 49 videos with complete rating data. All tests were two-sided, and a p value or adjusted p value <0.05, as applicable, was considered statistically significant.

Results

A total of 50 videos were included in the analysis. Of these, 37 (74%) were uploaded by personal/non-academic sources, 9 (18%) by academic associations or groups, and 4 (8%) by commercial/advertising sources. Symptoms of alcohol intoxication or poisoning were described in 30 videos (60%), whereas information on first intervention was provided in only 19 videos (38%). Three of the 50 included videos (6%) specifically addressed toxic alcohol poisoning; all focused-on methanol. None specifically addressed ethylene glycol or other toxic alcohols. Because of the small number of toxic-alcohol-related videos, a separate subgroup analysis or a sensitivity analysis was not considered statistically reliable. Complete DISCERN, GQS, and JAMA scores were available for 49 videos. The medians (IQRs) of the DISCERN, GQS, and JAMA scores were 27 (8), 3 (2), and 2 (1), respectively. The median video duration (IQR) was 3.78 (6.94) minutes; the median number of views was 19,964.50 (181,652.50), and the median number of likes was 188.00 (982.50). Comment counts were available for 49 videos because comments were disabled for one video; the median comment count (IQR) was 29.00 (325.50). The videos had a total duration of 285.98 minutes and accumulated 21,254,061 views, 259,796 likes, and 34,532 visible comments. The descriptive characteristics of the included videos are presented in Table 1.

Video characteristics by source category are presented in Table 2. When the videos were grouped by source, no statistically significant differences were found in DISCERN, GQS, or JAMA scores across sources classified as academic associations/groups, personal/non-academic, and commercial/advertising (p=0.207, p=0.271, and p=0.923, respectively). Significant overall differences were observed among video sources in view rate (p=0.003), video duration (p=0.002), view count (p=0.001), like count (p=0.001), and comment count (p=0.001). Dunn's post hoc tests with Bonferroni adjustment demonstrated that personal/non-academic videos had significantly higher view rates than academic videos (adjusted p=0.016) and commercial/advertising videos (adjusted p=0.049). Personal/non-academic videos were also significantly longer than academic (adjusted p=0.013) and commercial/advertising videos (adjusted p=0.039), and had significantly higher view counts (adjusted p=0.019 and p=0.007, respectively), like counts (adjusted p=0.016 and p=0.014, respectively), and comment counts (adjusted p=0.015 and p=0.009, respectively). No significant post-hoc differences were found between academic and commercial or advertising sources for these variables. The proportion of videos describing symptoms of intoxication or poisoning differed significantly across sources (exact p=0.007); all academic association/group videos described symptoms, compared with 48.6% of personal/non-academic videos and 75% of commercial/advertising videos. Pairwise Fisher's exact tests with Holm adjustment showed that symptom information was significantly more frequent in videos from academic associations/groups than in personal or non-academic videos (adjusted p=0.019). The other pairwise comparisons for symptom information were not statistically significant. Language format also differed significantly among video sources in the overall analysis (exact p=0.031); however, none of the pairwise comparisons for language format remained statistically significant after Holm adjustment. No significant source-based differences were found in country of upload, target audience, introduction of alcohol types, or first intervention information.

After Holm adjustment across the 15 unique correlation tests presented in Table 3, a strong positive correlation between DISCERN and GQS scores remained (rho=0.820, adjusted p<0.001). Moderate positive correlations also remained between DISCERN and JAMA scores (rho=0.455, adjusted p=0.014), between GQS and JAMA scores (rho=0.447, adjusted p=0.017), and between like rate and JAMA scores (rho=0.407, adjusted p=0.044). The correlations of like rate with DISCERN and with GQS, and of video duration with GQS, were not statistically significant after Holm adjustment. View rate and comment count were not significantly correlated with any of the three quality scores. The complete correlation results, including Holm-adjusted p values, are presented in Table 3.

Because the variables indicating the introduction of alcohol types, description of intoxication or poisoning symptoms, and

Table 1. Classification and descriptive characteristics of the included YouTube videos		
Characteristic		**Value**
Video duration, min, median (IQR)		3.78 (6.94)
Days since upload, median (IQR)		1655.50 (1878.00)
View count, median (IQR)		19,964.50 (181,652.50)
Like count, median (IQR)		188.00 (982.50)
Comment count, median (IQR)		29.00 (325.50)
Like rate, %, median (IQR)		0.79 (0.70)
View rate, median (IQR)		4108.49 (18,478.30)
Video source, n (%)	Academic association/group	9 (18)
	Personal/non-academic	37 (74)
	Commercial/advertisement	4 (8)
Country of upload, n (%)	United States	29 (58)
	Canada	4 (8)
	Other	7 (14)
	Unknown	10 (20)
Information introducing alcohol types, n (%)	No	47 (94)
	Yes	3 (6)
Description of alcohol intoxication/poisoning symptoms, n (%)	No	20 (40)
	Yes	30 (60)
Information on first intervention, n (%)	No	31 (62)
	Yes	19 (38)
Language format, n (%)	English audio	45 (90)
	English subtitles	5 (10)
Target audience, n (%)	Physicians	6 (12)
	Non-physicians/general public	44 (88)
DISCERN score, median (IQR)		27 (8)
GQS, median (IQR)		3 (2)
JAMA score, median (IQR)		2 (1)

Data are presented as median (IQR) or n (%). Comment count was available for 49 videos because comments were disabled for one video. Complete DISCERN, GQS, and JAMA scores were available for 49 videos. All three consensus scores were unavailable for one personal/non-academic video and could not be reconstructed because the video was no longer accessible during data verification. No rating scores were imputed. All other variables were available for all 50 videos. Like rate was calculated as (number of likes/number of views) × 100. View rate was calculated as (number of views × 365)/days since upload. DISCERN: Dissemination of Information on Consumer Health, GQS: Global Quality Score, JAMA: Journal of the American Medical Association, IQR, interquartile range

provision of first-intervention information were binary variables, they were not included in the correlation analysis. Instead, quality scores were compared between videos with and without each content characteristic using the Mann–Whitney U test. After Holm adjustment for multiple comparisons, videos describing intoxication or poisoning symptoms had significantly higher DISCERN scores than videos without symptom information [median (IQR): 30.50 (12.50) vs. 25.00 (4.00); adjusted p=0.002]. Videos providing first-intervention information also had higher DISCERN scores [36.00 (13.00) vs. 25.00 (4.00); adjusted p<0.001] and GQS scores [4.00 (1.00) vs. 2.00 (1.00); adjusted p=0.003]. JAMA scores did not differ significantly according to symptom content or first-intervention content. The introduction of alcohol types was not significantly associated with any quality score after adjustment for multiple comparisons (Table 4).

Discussion

YouTube as a Growing Health Information Platform

This study evaluated the first 50 YouTube videos related to alcohol intoxication and poisoning, which had accumulated more than 21 million views. In addition to the view count, the 259,796 likes and 34,532 visible comments further indicate substantial interest in the topic. Despite the potential benefits of platforms such as YouTube, these platforms may contribute to the promotion of substance use, particularly alcohol consumption, in contemporary society, partly because of the widespread accessibility and

Table 2. Video characteristics according to source category

Academic association/ group (n=9)		Video source			p value
		Personal/ non-academic (n=37)	Commercial/ advertisement (n=4)		
DISCERN score (median (IQR)		31 (12)	27.50 (8)	25.50 (4.75)	0.207
GQS (median (IQR)		3 (1.50)	3 (2)	2 (0.75)	0.271
JAMA score (median (IQR)		2 (1)	2 (1)	2 (0.75)	0.923
Like rate (%) (median (IQR)		0.66 (0.59)	0.99 (0.79)	0.64 (0.22)	0.075
View rate (median (IQR)		736.31 (4959.69)	6261.72 (137,385.66)	867.19 (1249.16)	**0.003**
Video duration, min (median (IQR)		1.42 (1.41)	5.45 (8.48)	1.43 (0.98)	**0.002**
Days since upload (median (IQR)		1961 (1470)	1790 (2123)	1028 (1300.50)	0.285
View count (median (IQR)		3506 (30,652)	32,852 (563,745.50)	1554.50 (1554.25)	**0.001**
Like count (median (IQR)		15 (273)	298 (6599.50)	8 (10.25)	**0.001**
Comment count (median (IQR)		0.50 (38.50)	50 (738.50)	0 (0.75)	**0.001**
Country of upload	USA (n; %)	5 (55.6)	20 (54.1)	4 (100)	0.924
	Canada (n; %)	1 (11.1)	3 (8.1)	0 (0)	
	Other (n; %)	1 (11.1)	6 (16.2)	0 (0)	
	Unknown (n; %)	2 (22.2)	8 (21.6)	0 (0)	
Information introducing alcohol types	No (n; %)	9 (100)	34 (91.9)	4 (100)	1.000
	Yes (n; %)	0 (0)	3 (8.1)	0 (0)	
Description of intoxication/poisoning symptoms	No (n; %)	0 (0)	19 (51.4)	1 (25)	**0.007**
	Yes (n; %)	9 (100)	18 (48.6)	3 (75)	
Information on first intervention	No (n; %)	5 (55.6)	24 (64.9)	2 (50)	0.694
	Yes (n; %)	4 (44.4)	13 (35.1)	2 (50)	
Language format	English audio (n; %)	8 (88.9)	35 (94.6)	2 (50)	**0.031**
	English subtitles (n; %)	1 (11.1)	2 (5.4)	2 (50)	
Target audience	Physicians (n; %)	1 (11.1)	5 (13.5)	0 (0)	1.000
	Non-physicians/general public (n; %)	8 (88.9)	32 (86.5)	4 (100)	

Data are presented as median (IQR) or n (%). Continuous variables were compared using the Kruskal-Wallis test. When the overall Kruskal-Wallis test was statistically significant, Dunn's pairwise post hoc test with Bonferroni adjustment was performed. Categorical variables were compared using the Fisher-Freeman-Halton exact test because of small expected cell counts. Following statistically significant Fisher-Freeman-Halton omnibus tests, pairwise two-sided Fisher's exact tests were performed with Holm adjustment for multiple comparisons. After adjustment, the difference in symptom information remained statistically significant between academic association/group and personal/non-academic sources (Holm-adjusted p=0.019), whereas none of the pairwise comparisons for language format remained statistically significant. Rating data were available for 49 videos (academic association/group, n=9; personal/non-academic, n=36; commercial/advertisement, n=4). All three quality scores (DISCERN, GQS, and JAMA) were unavailable for one personal/non-academic video, and no missing scores were imputed. Comment count was available for 49 videos (academic association/group, n=8; personal/non-academic, n=37; commercial/advertisement, n=4). DISCERN: Dissemination of Information on Consumer Health, GQS: Global Quality Score, JAMA: Journal of the American Medical Association, IQR: Interquartile range, USA: United States of America

influence of social media (1,2). The use of YouTube as a medical educational resource among medical students, who will become future academicians and clinicians, has reached rates of 82.3%-91.8% (19-21). Furthermore, the number of ED visits related to alcohol consumption is increasing (5-7). Considering these factors, the viewership of videos related to alcohol intoxication among healthcare professionals is expected to continue to rise. YouTube surpasses traditional text-based resources by offering visual and auditory content, making it an attractive tool for disseminating medical information (21). It also shows that not only healthcare professionals but also patients and their families use YouTube as an information source in emergency situations, preoperative contexts, and non-emergency settings (22,23).

Risks of Misinformation and the Role of Unregulated Content

Although YouTube offers potential for public health education, it also presents significant risks. The absence of an editorial review process means that videos may contain inaccurate or misleading information. Viewers often cannot assess the credibility of sources,

Table 3. Spearman correlations between video quality scores and quantitative or ordinal video characteristics			
Variable	**DISCERN, rho (Holm-adjusted p value)**	**GQS, rho (Holm-adjusted p value)**	**JAMA, rho (Holm-adjusted p value)**
GQS	0.820 (<0.001)	—	—
JAMA	0.455 (0.014)	0.447 (0.017)	—
Like rate	0.339 (0.156)	0.381 (0.073)	0.407 (0.044)
View rate	0.041 (1.000)	0.231 (0.886)	0.005 (1.000)
Video duration	0.178 (1.000)	0.383 (0.073)	0.225 (0.886)
Comment count	-0.029 (1.000)	0.073 (1.000)	0.061 (1.000)

Spearman's rank correlation analysis was restricted to quantitative and ordinal variables. The Holm procedure was applied across the 15 unique correlation tests presented in the table, and Holm-adjusted two-sided p values are reported. Correlations involving DISCERN, GQS, or JAMA scores were based on 49 videos because all three rating scores were unavailable for one personal/non-academic video. Correlations involving comment count and a quality score were based on 48 videos because comments were disabled for one video, while the three quality scores were unavailable for a different video. No missing values were imputed. Binary and non-ordinal nominal variables were not included in the correlation analysis. DISCERN: Dissemination of Information on Consumer Health, GQS: Global Quality Score, JAMA: Journal of the American Medical Association, IQR: Interquartile range, USA: United States of America

Table 4. Comparison of video quality scores according to the presence of specific content				
Content characteristic	**Score**	**Absent, median (IQR)**	**Present, median (IQR)**	**Holm-adjusted p value**
Introduction of alcohol types	DISCERN	26.50 (8.00)	38.00 (10.00)	0.378
	GQS	3.00 (2.00)	4.00 (0.00)	0.273
	JAMA	2.00 (1.00)	2.00 (1.00)	0.981
Description of intoxication/poisoning symptoms	DISCERN	25.00 (4.00)	30.50 (12.50)	**0.002**
	GQS	3.00 (1.00)	3.00 (2.00)	0.353
	JAMA	2.00 (1.00)	2.00 (1.00)	0.518
Information on first intervention	DISCERN	25.00 (4.00)	36.00 (13.00)	**<0.001**
	GQS	2.00 (1.00)	4.00 (1.00)	**0.003**
	JAMA	2.00 (1.00)	2.00 (1.00)	0.714

Comparisons were performed using the Mann-Whitney U test. P values were adjusted across the nine comparisons using the Holm procedure. Rating data were available for 49 videos because DISCERN, GQS, and JAMA scores were all unavailable for one personal/non-academic video; no missing scores were imputed. For the introduction of alcohol types, the absent and present groups included 46 and 3 videos, respectively. For symptom information, the absent and present groups included 19 and 30 videos, respectively. For first-intervention information, the absent and present groups included 30 and 19 videos, respectively. DISCERN: Dissemination of Information on Consumer Health, GQS: Global Quality Score, JAMA: Journal of the American Medical Association, IQR: Interquartile range, USA: United States of America

and exposure to deceptive advertising further complicates the situation (23,24). Studies examining the scientific presentation quality of YouTube content in various health-related fields have highlighted that video content often lacks high-quality information, which is considered a significant limitation of the platform (23,25,26). Indeed, the findings that 8% of videos in the present study contained commercial/advertisement content and that only 60% described signs or symptoms of alcohol-related intoxication or poisoning are consistent with this concern.

Quality Scores According to Uploader Source

No statistically significant differences in DISCERN, GQS, or JAMA scores were detected according to uploader source, contrary to some previous studies reporting higher quality for videos uploaded by academic institutions or healthcare professionals (27-29). However, the absence of statistically significant differences should not be interpreted as evidence that the source categories were equivalent. The small and unequal group sizes, particularly the inclusion of only four videos in the commercial/advertisement category, limited the statistical power and precision of source-based comparisons. Instead, evaluation of the actual content remains essential. This implies that efforts to address alcohol-related misinformation on social media must involve not only institutions but also individuals, through cooperative public health strategies. The content-based comparisons further clarify this finding. Videos describing the signs or symptoms of alcohol-related intoxication or poisoning had significantly higher DISCERN scores than videos without symptom information. Importantly, videos providing first-intervention information had substantially higher DISCERN and GQS scores than those without such information. These findings support the relevance of clinically actionable content, but do not establish that the uploader source is unrelated to video quality. However, these associations should not be interpreted as causal. The inclusion of information on symptoms or interventions may directly contribute to higher DISCERN and GQS scores because these instruments assess the completeness, usefulness, and educational quality of the information presented.

In contrast, JAMA scores did not differ significantly based on the presence of symptom information or first-intervention information. This apparent difference among the scoring instruments is consistent with their distinct purposes. DISCERN and GQS evaluate the quality, usefulness, and presentation of clinical information, whereas the JAMA benchmark criteria primarily assess authorship, attribution, disclosure, and currency. Consequently, a video may provide clinically useful information while still receiving a relatively low JAMA score if its sources, authorship, or update information are not adequately reported. Similarly, videos introducing specific types of alcohol did not have significantly different quality scores after adjustment for multiple comparisons. Because only three videos included information on alcohol types, this finding should be interpreted with caution and not considered evidence that alcohol-type information is unrelated to video quality.

Clinical Content Gaps in Alcohol-related Videos

The present study identified important deficiencies in the clinical content of the evaluated videos. Most videos (94%) did not provide information distinguishing among alcohol types; 62% did not include any actionable information on initial intervention. Although 60% described signs or symptoms of alcohol-related intoxication or poisoning, symptom recognition, without clear guidance on appropriate immediate actions, may have limited practical value for viewers facing an actual emergency. These findings indicate that many highly accessible videos provide only partial information and may not adequately link recognition of a potentially serious condition to the need for timely and appropriate medical assistance.

Previous studies have reported that alcohol-related content on YouTube may emphasize socially engaging, favorable, or humorous portrayals while giving less attention to dependence, withdrawal, and other adverse health consequences (13). However, the present study did not include a predefined variable to assess whether videos were humorous, entertainment-oriented, favorable toward alcohol use, or potentially glamorizing. Therefore, no direct conclusions can be drawn from the current data regarding the tone or entertainment orientation of the included videos. The findings of the present study should instead be interpreted as demonstrating measurable gaps in clinical content, particularly in the presentation of alcohol types and in actionable information for first interventions.

Emergency Medicine Relevance of First-intervention Content

The limited availability of first-intervention information is particularly important from the perspective of emergency medicine. Although 60% of the evaluated videos described signs or symptoms of alcohol-related intoxication or poisoning, only 38% provided any actionable information on immediate response or early medical management. This finding indicates a substantial gap between symptom recognition and guidance on what viewers should do when confronted with a potentially life-threatening alcohol-related emergency. Importantly, in the present study, the presence of first-intervention information was recorded as a binary content variable and should not be taken to indicate that the information was complete, fully accurate, or concordant with all relevant recommendations.

Emergency medicine guidance emphasizes that the initial management of acute ethanol intoxication should prioritize assessment and stabilization of airway, breathing, and circulation; identification and treatment of hypoglycemia; protection against aspiration; evaluation for trauma, co-ingestion, and alternative causes of altered mental status; and appropriate monitoring and supportive care (3,4,12). In individuals at risk of nutritional deficiency or chronic alcohol use, thiamine administration should also be considered within the clinical management pathway (3,12). Although these recommendations are primarily intended for healthcare professionals, public-facing videos should translate the most time-sensitive principles into clear and safe actions, such as activating emergency medical services, avoiding potentially harmful practices, monitoring consciousness and breathing, and not leaving a person with impaired consciousness unattended.

The distinction between acute ethanol intoxication and toxic alcohol poisoning is also clinically important. Methanol and ethylene glycol poisoning may initially present with nonspecific features but can progress to severe metabolic acidosis, visual impairment, neurological deterioration, renal failure, or death. Their management may require urgent antidotal treatment and enhanced elimination techniques, including hemodialysis (8-10). Therefore, videos addressing toxic alcohol exposure should not present such cases as equivalent to uncomplicated ethanol intoxication or imply that observation alone is sufficient. Because ethanol-related and toxic-alcohol-related videos were not classified as separate analytical groups in the present study, the adequacy of intervention content for each specific toxicological condition could not be compared.

The low proportion of videos containing actionable first-intervention information may have clinically relevant consequences. Individuals and relatives may consult online videos before contacting emergency medical services; incomplete or misleading guidance may delay emergency evaluation or encourage inappropriate responses. From an emergency medicine standpoint, high-quality public-facing content should therefore combine symptom recognition with explicit recommendations regarding when and how to seek

urgent medical assistance. Collaboration among emergency physicians, toxicologists, professional organizations, and digital platforms may help ensure that highly visible videos provide concise, clinically appropriate, and actionable information for alcohol-related emergencies.

Study Limitations

This study has several limitations. First, it is a cross-sectional assessment of the first 50 eligible YouTube videos identified using two specific search terms on a single date. YouTube search rankings are dynamic and may change according to platform algorithms, geographic location, user behavior, and changes in video popularity. Although an incognito browser session was used and no sorting or geographic filter was applied, the influence of all platform-level personalization and ranking mechanisms could not be completely eliminated. Furthermore, because the data were collected on October 1, 2023, subsequent changes in video availability, content, ranking, and engagement metrics are not reflected in the analysis. Therefore, the findings should be interpreted as a snapshot of the platform at the time of data collection rather than as a permanent characterization of YouTube content.

Second, the analysis was limited to videos available in English through either English-language audio or English subtitles. Relevant content in other languages may differ in quality, clinical emphasis, and recommendations for emergency care. The study was also restricted to YouTube and did not evaluate other widely used platforms, including Instagram, TikTok, Facebook, X, or Snapchat. Consequently, the findings may not be generalizable to other languages, platforms, search strategies, or social media environments.

Third, all videos were evaluated simultaneously by two experienced emergency medicine physicians using a consensus-based approach. Because independent ratings were not recorded, inter-rater reliability could not be quantified using an intraclass correlation coefficient or another agreement statistic. Although consensus assessment may reduce unresolved disagreement, it does not demonstrate the reproducibility of the scores across independent evaluators. In addition, the evaluators were not blinded to visible video characteristics, including uploader identity and engagement metrics, because these characteristics formed part of the study data and were publicly displayed on the platform. This may have introduced evaluator-related bias.

Fourth, the variables assessing symptom and first-intervention information were recorded as binary indicators of the presence of content. These variables did not measure the completeness, clinical accuracy, comprehensibility, or full concordance of the information provided with clinical guidelines. A video classified as containing first-intervention information may, therefore, have included only one relevant recommendation and should not necessarily be considered a complete or fully guideline-concordant educational resource. Similarly, the study did not use a predefined instrument to evaluate potentially harmful recommendations, misinformation severity, viewer comprehension, or the tone and entertainment orientation of the videos.

Fifth, the search strategy primarily targeted acute ethanol intoxication. Only three videos specifically addressed toxic alcohol poisoning; all focused-on methanol, and none addressed ethylene glycol or any other toxic alcohol. Because of the small number of toxic-alcohol-related videos, a separate subgroup or sensitivity analysis was not feasible. Acute ethanol intoxication and toxic alcohol poisoning differ in clinical presentation, urgency, and treatment; therefore, including both conditions in the overall sample may have introduced clinical heterogeneity and limited condition-specific interpretation of findings from the first intervention.

Finally, some subgroup analyses were based on small sample sizes, particularly the commercial/advertising category and videos introducing alcohol types. This reduced statistical power and limited the precision of comparisons involving these categories. Rating data were unavailable for one video, and the comment count was unavailable for another video because comments were disabled; these observations were handled using an available-case approach without imputation. Engagement measures, such as views, likes, and comments, are influenced by video age and platform exposure, and should not be interpreted as direct indicators of clinical quality. Because of the observational and cross-sectional design, the identified associations cannot establish causal relationships between video content, uploader characteristics, engagement, and quality scores.

Conclusion

YouTube videos addressing alcohol-related intoxication and poisoning attract substantial public attention, but frequently provide incomplete, clinically actionable information. Although 60% of the evaluated videos described signs or symptoms, only 38% provided information on first intervention. No statistically significant differences in video quality or reliability scores were detected across uploader sources. However, this finding should not be interpreted as evidence of equivalence among source categories, particularly given the small and unequal group sizes and the inclusion of only four commercial/advertisement videos. In contrast, videos containing symptom information had higher DISCERN scores, while those providing first-intervention information had higher DISCERN and GQS scores. These findings

support the relevance of clinically actionable content, but do not establish equivalence among uploader source categories or demonstrate that the uploader source is unrelated to video quality.

Public-facing content should clearly distinguish acute ethanol intoxication from toxic alcohol poisoning caused by substances such as methanol or ethylene glycol and should combine symptom recognition with explicit, safe, and timely recommendations for seeking emergency medical assistance. Greater involvement of emergency physicians, toxicologists, and professional organizations, together with transparent authorship, appropriate source attribution, periodic content review, and stronger platform-level editorial oversight, may improve the quality and clinical usefulness of online information on alcohol-related emergencies. Viewers and healthcare professionals should remain cautious when using uploader's identity, popularity, or engagement metrics as proxies for the reliability of medical information.

Ethics

Ethics Committee Approval: Ethics committee approval was not required because this observational, cross-sectional study analyzed publicly available YouTube videos and did not involve human participants, animal subjects, interventions, or participant interaction.

Informed Consent: Not applicable.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Footnotes

Authorship Contributions

Surgical and Medical Practices: H.Y.B., S.A., Concept: H.Y.B., S.A., Design: H.Y.B., S.A., Data Collection or Processing: H.Y.B., S.A., Analysis or Interpretation: H.Y.B., S.A., Literature Search: H.Y.B., S.A., Writing: H.Y.B., S.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Niu S, Reed KP. Towards Understanding substance abuse misinformation in YouTube Videos. In: Computer Supported Cooperative Work and Social Computing (CSCW '23 Companion); October 14-18, 2023; Minneapolis, MN, USA. New York, NY: ACM; 2023.
2. Merrill JE, López G, Doucette H, Pielech M, Corcoran E, Egbert A, et al. Adolescents' perceptions of alcohol portrayals in the media and their impact on cognitions and behaviors. Psychol Addict Behav. 2023;37:758-70.
3. Strayer RJ, Friedman BW, Haroz R, Ketcham E, Klein L, LaPietra AM, et al. Emergency department management of patients with alcohol intoxication, alcohol withdrawal, and alcohol use disorder: a white paper prepared for the American Academy of Emergency Medicine. J Emerg Med. 2023;64:517-40.
4. Mirijello A, Sestito L, Antonelli M, Gasbarrini A, Addolorato G. Identification and management of acute alcohol intoxication. Eur J Intern Med. 2023;108:1-8.
5. Hanrahan M, O'Mahony M, McLoughlin D, Sheahan A. Examination of patients presenting to the emergency department as apparently drunk. Ir J Med Sci. 2025;194:375-84.
6. Maharaj T, Fitzgerald N, Gilligan E, Quirke M, MacHale S, Ryan JD. Alcohol-related emergency department presentations and hospital admissions around the time of minimum unit pricing in Ireland. Public Health. 2024;227:38-41.
7. White AM, Slater ME, Ng G, Hingson R, Breslow R. Trends in alcohol-related emergency department visits in the United States: results from the nationwide emergency department sample, 2006 to 2014. Alcohol Clin Exp Res. 2018;42:352-9.
8. Moral AR, Cankayali I, Sergin D, Boyacilar O. Neuromuscular functions on experimental acute methanol intoxication. Turk J Anaesthesiol Reanim. 2015;43:337-43.
9. Ng PCY, Long BJ, Davis WT, Sessions DJ, Koyfman A. Toxic alcohol diagnosis and management: an emergency medicine review. Intern Emerg Med. 2018;13:375-83.
10. Paasma R, Hovda KE, Jacobsen D. Methanol poisoning and long term sequelae - a six years follow-up after a large methanol outbreak. BMC Clin Pharmacol. 2009;9:5.
11. LaHood AJ, Kok SJ. Ethanol toxicity. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2023.
12. Sarkar S, Bhatia G, Dhawan A. Clinical practice guidelines for assessment and management of patients with substance intoxication presenting to the emergency department. Indian J Psychiatry. 2023;65:196-211.
13. Primack BA, Colditz JB, Pang KC, Jackson KM. Portrayal of alcohol intoxication on YouTube. Alcohol Clin Exp Res. 2015;39:496-503.
14. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor--Let the reader and viewer beware. JAMA. 1997;277:1244-5.
15. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis - a wakeup call? J Rheumatol. 2012;39:899-903.
16. Erdem MN, Karaca S. Evaluating the accuracy and quality of the information in kyphosis videos shared on YouTube. Spine (Phila Pa 1976). 2018;43:E1334-E9.
17. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: An instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health. 1999;53:105-11.
18. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg. 2018;126:1763-8.
19. Alzoubi H, Karasneh R, Irshaidat S, Abuelhaija Y, Abuorouq S, Omeish H, et al. Exploring the use of YouTube as a pathology learning tool and its relationship with pathology scores among medical students: cross-sectional study. JMIR Med Educ. 2023;9:e45372
20. Pradhan S, Das C, Panda DK, Mohanty BB. Assessing the utilization and effectiveness of YouTube in anatomy education among medical students: a survey-based study. Cureus. 2024;16:e55644.

21. Akakpo MG, Akakpo PK. Recognizing the role of YouTube in medical education. Discov Educ. 2024;3:73.
22. Toksoz A, Duran MB. Analysis of videos about vesicoureteral reflux on YouTube. J Pediatr Urol. 2021;17:858.e1-858.e6.
23. Roberts B, Kobritz M, Nofi C, Demyan L, Guevara J, Hansen L, et al. Social media, misinformation, and online patient education in emergency general surgical procedures. J Surg Res. 2023;287:16-23.
24. Staunton PF, Baker JF, Green J, Devitt A. Online curves: a quality analysis of scoliosis videos on YouTube. Spine (Phila Pa 1976). 2015;40:1857-61.
25. Evran T, Ilhan S. Anesthesia videos in geriatric and elderly patients on YouTube: content, quality, reliability, and usefulness assessment. PeerJ. 2025;13:e19280.
26. Temel MH, Erden Y, Bagcier F. Assessing patients perception: analyzing the quality, reliability, comprehensibility, and the mentioned medical concepts of traumatic brain injury videos on YouTube. World Neurosurg. 2024;185:e907-e14.
27. Sasmaz MI, Akca AH. Reliability of trauma management videos on YouTube and their compliance with ATLS® (9th edition) guideline. Eur J Trauma Emerg Surg. 2018;44:753-7.
28. Azer SA, AlGrain HA, AlKhelaif RA, AlEshaiwi SM. Evaluation of the educational value of YouTube videos about physical examination of the cardiovascular and respiratory systems. J Med Internet Res. 2013;15:e241.
29. Lee JS, Seo HS, Hong TH. YouTube as a potential training method for laparoscopic cholecystectomy. Ann Surg Treat Res. 2015;89:92.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.65037
Eurasian J Emerg Med. 2026;25: 449-55

Clinical and Endoscopic Features of Alkaline Toxicity in Children: A 5-year Retrospective Cohort Study

Shafiqa Alsharif

King Abdulaziz University Faculty of Medicine, Department of Emergency Medicine, Jeddah, Saudi Arabia

Abstract

Aim: Alkaline toxicity is a rare emergency condition that may result in substantial health issues in children with long-term morbidities. Our study aims to describe the clinical and endoscopic features of alkaline ingestion in children and to review the management.

Materials and Methods: A retrospective cohort analysis was conducted on 33 pediatric patients, who ingested weak or strong alkaline substance and presented to the emergency department in a single tertiary center.

Results: The mean patients' age was 24±18.03 months with 17 (51.5%) boys and 16 (48.5%) girls. Twenty cases (60.6%) ingested strongly alkaline substances, and 13 cases (39.4%) ingested weakly alkaline substances. The common presenting symptoms were vomiting (63.6%), drooling (39.4%), hematemesis (12.1%), and stridor and chest wall burns (6.1%). Shortness of breath and oral stenosis were rare (3%). Esophagogastroduodenoscopy (EGD) was performed within 48 hours on children who had ingested a strong alkali. The endoscopic findings followed Zargar's classification. Suicide attempts or deaths were not reported in any cases. Clinical manifestations did not differ significantly between the exposure groups; however, endoscopic findings were not compared because EGD was performed only in children who had ingested strong alkali. Among the 20 children who underwent EGD, esophageal burns were identified in 8 (40%).

Conclusion: Our findings highlight the role of early endoscopic evaluation in identifying gastrointestinal injury after alkaline ingestion and guiding subsequent clinical management in affected children.

Keywords: Alkaline ingestion, clinical features, complications, endoscopy, children

Introduction

Unintentional ingestion of corrosives has been addressed in multiple guidelines and protocols over the past several years, particularly in the pediatric population. Several studies and case reports have demonstrated the catastrophic events that result immediately following the ingestion and over the long-term follow-ups (1). Morbidity outcomes vary depending on the type and quantity of corrosive ingested. The acidity of substances is determined by their pH. Substances with a pH above 7 are classified as alkaline, with those exceeding a pH of 11 requiring special attention due to potential human exposure (2). Alkaline toxicity is a rare health issue among adults and children; children have a higher incidence of alkaline than acidic ingestion. Males tend to have a higher frequency of ingesting alkaline substances than females (3,4). Common alkaline substances include bleach, sodium hydroxide, potassium hydroxide, ammonia, and dishwashing detergents (5). Accidental ingestion of small amounts of these substances typically results in minimal health effects. However, intentional ingestion, or ingestion in large quantities, can lead to severe consequences, such as injuries to the eyes, skin, esophagus, and gastrointestinal tract (GIT). Those patients require comprehensive medical intervention and prolonged treatment.

In developed countries, the morbidity and mortality rates related to caustic ingestion have decreased due to stricter product regulations and increased public awareness about alkaline products (6). Despite these efforts, caustic substances continue to prompt numerous calls to control centers and visits to emergency departments (7). In Saudi Arabia, morbidity from caustic esophageal burns remains a significant health concern. The accessibility of these detergents varies by location, although

Corresponding Author: Asst. Prof. Shafiqa Alsharif, King Abdulaziz University Faculty of Medicine, Department of Emergency Medicine, Jeddah, Saudi Arabia
E-mail: halshrif@kau.edu.sa **ORCID ID:** orcid.org/0009-0006-4744-3599

Cite this article as: Alsharif S. Clinical and endoscopic features of alkaline toxicity in children: a 5-year retrospective cohort study. Eurasian J Emerg Med. 2026;25: 449-55

Received: 16.05.2026
Accepted: 05.09.2026
Published: 02.10.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

the manufacturers' containers are securely sealed. The severity of damage resulting from alkaline ingestion is influenced by the duration of contact with the affected surface, the amount ingested, and the titratable alkali reserve (TAR) (8). TAR represents the quantity of acid needed to neutralize an alkaline substance to a pH of 8.0. However, this information is often unavailable during emergencies. Alkaline substances with a pH above 7 can cause liquefactive necrosis, while those with a pH exceeding 11 are likely to result in severe injuries (9). Alkaline ingestion may cause esophageal damage including necrosis, proliferation, and stricture formation (4,10). It can also be reactive with the stomach hydrochloride causing alkaline gas that leads to gastric necrosis, and respiratory irritation (10). Direct contact with the alkaline substance can also result in skin burn and corneal injury (11). Patients who ingest alkaline substances may present with several symptoms including acute abdominal pain, vomiting, hematemesis, dysphagia, and rarely chest pain and shortness of breath (12). Stridor and saliva drooling are also reported symptoms (13). Patients presenting with alkaline ingestion, regardless of the amount taken, require proper care and management, starting with rapid assessment of the affected area, decontamination by flushing with copious amounts of water, and prompt medical attention. Blood tests may show leukocytosis, coagulopathy, lactic acidosis, or, rarely, renal or hepatic failure if the toxicity is severe. Radiological imaging may not help but sometimes can rule out a perforated viscus (9). Esophagogastroduodenoscopy (EGD) should be considered in children with suspected caustic ingestion who develop symptoms or signs suggestive of gastrointestinal injury, including vomiting, drooling, dysphagia, hematemesis, respiratory symptoms, abdominal pain, or oral lesions. In asymptomatic children without oral lesions, endoscopy may be withheld, provided that appropriate clinical observation and follow-up are ensured (1-3). The assessment of the esophagus in cases of alkali ingestion followed Zargar classification system (14).

Treatment always starts with airway protection (14). The placement of a nasogastric tube (NG tube) or the use of steroids in cases of caustic ingestion is still controversial. However, dexamethasone may reduce the airway inflammation and edema (15,16). Pain management and nutritional support are crucial, particularly for children and in long-term care. Generally, children with grade 1 injuries can safely be discharged home once they tolerate oral intake. Grade 2a injuries require close monitoring in the hospital until there is clinical improvement. Grade 2b/3 injuries necessitate admission to the intensive care unit (ICU) for appropriate management. Grades 2b and 3 burns have the potential to develop dense scar tissue and stricture formation, requiring long-term care involving dilations or stenting to maintain esophageal patency (17). Squamous cell carcinoma of the esophagus can be a complication of grade 3 injuries (18).

Noting numerous children presenting to the pediatric emergency department with acute complications and occasionally, irreversible damage to multiple organ systems, we conducted a retrospective cohort study to examine these cases in terms of common presentations, grading features, and clinical management.

Materials and Methods

This retrospective cohort study included pediatric patients who presented to the Emergency Department of King Abdulaziz University Hospital following alkaline substance ingestion between 2012 and 2017. Patients were eligible if they were younger than 18 years, had a documented history of alkaline ingestion, and had sufficient clinical information in the medical record to confirm the exposure and initial clinical presentation. Patients who ingested non-alkaline corrosive substances or who lacked sufficient documentation to confirm alkaline exposure were excluded. Informed consent was obtained from the parents before performing the procedure. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Biomedical Ethics Committee of King Abdulaziz University, Faculty of Medicine (approval no: 483-16. date: January 8, 2017). Informed consent was obtained from all parents before the procedure and the study.

Eligible cases were identified retrospectively from hospital medical records of children presenting to the emergency department with documented alkaline ingestion during the study period. Demographic characteristics, exposure history, presenting symptoms and signs, type and physical form of the alkaline substance, endoscopic findings, treatment, complications, and available follow-up information were extracted from the medical records. The 33 patients enrolled in this study were children <18 years of age and of both genders. All patients were evaluated at presentation by both emergency physicians and gastroenterologists. Clinical manifestations and physical examinations of the heart, lungs, and abdomen were performed. EGD was performed on 20 children with documented ingestion of a strongly alkaline substance who were clinically stable to undergo the procedure. Children with exposure to weak alkali and mild clinical manifestations were managed conservatively and did not routinely undergo EGD. Consequently, endoscopic outcomes were analyzed descriptively only among patients who underwent EGD; the outcomes were not statistically compared between the weak- and strong-alkali groups. Clinical information and patients' data used in this study, including age, sex, symptoms, history of ingestion, and the type of alkali (weak

or strong; fluid or granular), were retrieved from the hospital records.

Alkaline exposures were categorized as weak or strong according to the reported pH of the ingested substances. Substances with a pH >7 and ≤11 were classified as weak alkalis, whereas substances with a pH >11 were classified as strong alkalis. When available, the specific product and its documented composition or concentration were reviewed to support this classification. Based on the records, the endoscopic examination was performed under local anesthesia in those who had ingested a strong alkaline substance within 48h of admission.

Statistical Analysis

Given the small sample size and low frequency of several clinical events, the statistical analysis was primarily descriptive and exploratory. Data were described as values and percentages, including the mean and standard deviation (SD). Comparisons between weak- and strong-alkali groups were restricted to clinical variables assessed in both groups. Endoscopy-dependent findings were summarized for only the 20 patients who underwent EGD; no between-group hypothesis testing was performed. Fisher's exact test was used to compare categorical data among the study groups using Statistical Package for Social Sciences (SPSS) software version 22 (SPSS Inc., Chicago, IL, USA).

A p value<0.05 was considered statistically significant.

Results

A total of 33 children (age, mean ± SD: 24±18.03 months; Range: 12-96 months); 17 (51%) males and 16 (48.5%) females, were included in this study. All patients presented to the emergency department with ingestion of strong (n=20) or weak (n=13) alkaline solutions. All patients were admitted and managed appropriately during the study period. Accidental ingestion was documented in 19 patients (57.6%), while information regarding the intent of ingestion was unavailable for the remaining 14 patients (42.4%). No suicide attempts were reported (Tables 1, 2).

The number of substances ingested and the time elapsed after ingestion are summarized in Table 2. Presenting symptoms were evaluated in all 33 patients. Vomiting was one of the most frequently documented symptoms, occurring in 21 patients (63.6%) (Table 3).

As outlined in Table 4, ingestion of strong alkali often presented with symptoms such as vomiting, drooling, oral edema, and chest stridor, whereas ingestion of weak alkali showed no significant symptoms except mild vomiting and oral cutaneous edema. Because the majority of patients who ingested weak alkaline substances experienced minimal, non-serious symptoms, EGD was performed only on patients who had ingested strong alkaline substances (n=20; 60.6%).

Ingestion of strongly alkaline substances was also associated with extensive mucosal involvement and complications. Clinical manifestations that were assessed in both exposure groups were compared using Fisher's exact test. Endoscopic findings were not compared between the weak- and strong-alkali groups because EGD was performed only in the strong-alkali group. Among the 20 children who underwent EGD, esophageal burns were identified in 8 (40%) children (Table 4). However, esophageal strictures and perforations were observed in 20% and 5% of patients who underwent endoscopy, respectively (Figure 1).

Those patients each required esophageal balloon dilatation under general anesthesia more than five times (Figure 1). Pharyngeal and gastric erosions were each documented in 4

Table 1. Demographic data of the total patients included in the study (n=33)

Age at baseline, months (mean ± SD)	**24±18.03**
Age at baseline, months (range)	12-96
Age at baseline, months n (%)	
12	14 (42.4%)
13	0 (0%)
18	3 (9.1%)
19	0 (0%)
24	7 (21.1%)
36	5 (15.2%)
48	2 (6.1%)
60	1 (3%)
96	1 (3%)
Sex, n (%)	
Male	17 (51.5%)
Female	16 (48.5%)

Table 2. Data shows the alkali amount ingested the time elapsed after ingestion all children involved in the study (available data, n=21)

Substance	**n (%)**
Weak alkali	13 (39.4%)
Strong alkali	20 (60.6%)
Time elapsed since ingestion	**n (%)***
Minutes	4 (19%)
1 hour	8 (38.1%)
2 hours	2 (9.5%)
3 hours	2 (9.5%)
>24 hours	5 (23.8%)

*Available data, n=21

of the 20 patients who ingested a strongly alkaline substance (20.0%), Gastrostomy tube placement was required in 2 of 20 patients who ingested a strongly alkaline substance (10.0%).

Additional complications resulting from the ingestion of strong alkaline substances and from severe morbidities are outlined in Table 4. While no fatalities were reported among the patients, three individuals required admission to the pediatric ICU for mechanical ventilation; this subsequently led to complications, including mediastinitis and airway fistula.

Discussion

The incidence of corrosives substances ingestion, including acids and alkalis, has decreased in developed countries due to safety precautions levied on hazardous products (19), while it is still a major burden in developing countries where the living environment can be unsafe. Nevertheless, some developed countries continue to face challenges, including a few cases that have not been reported in the literature. Our study focused on Saudi children under the age of 14 who presented to the emergency department of our hospital following alkaline ingestion. Accidental ingestion was documented in 57.6% of patients; however, information regarding the intent of ingestion was unavailable for 42.4% of patients. Accidental ingestion implies that children were not adequately prevented from accessing locations where these corrosive substances were stored. Rafeey et al. (19) observed that children who were just beginning to walk or crawl were particularly susceptible and experienced significant morbidity and mortality, with the highest

Table 3. Common presentations of patients ingested weak or strong alkalis

Symptom at presentation	Total (n=33)	Weak alkali (n=13)	Strong alkali (n=20)
Vomiting	21 (63.6%)	3 (23.1%)	18 (90%)
Cyanosis	1 (3%)	0 (0%)	1 (5%)
Shortness of breath	1 (3%)	0 (0%)	1 (5%)
Saliva drooling	13 (39.4%)	0 (0%)	13 (65%)
Hematemesis	4 (12.1%)	0 (0%)	4 (20%)
Oral edema	21 (63.6%)	3 (23.1%)	18 (90%)
Stridor	2 (6.1%)	0 (0%)	2 (10%)

Percentages in the total column were calculated using the overall cohort denominator (n=33), whereas percentages in the weak- and strong-alkali columns were calculated using their respective subgroup denominators (n=13 and n=20)

Table 4. Morbidities and complications according to the type of alkaline substance ingested

		Weak alkali (n=13)	Strong alkali (n=20)
ENT system	Burns of oropharynx	0 (0%)	4 (20%)
	Air way edema	0 (0%)	3 (15%)
	Tracheostomy	0 (0%)	1 (5%)
Respiratory system	Mechanical ventilation	0 (0%)	2 (10%)
	Air way fistula	0 (0%)	1 (5%)
	Mediastinitis	0 (0%)	1 (5%)
Gastrointestinal system	Oral edema	0 (0%)	5 (25%)
	Tongue edema	0 (0%)	3 (15%)
	Pharyngeal erosions	Not assessed	4 (20%)
	Esophageal burns	Not assessed	**8 (40%)**
	Gastric erosions	Not assessed	4 (20%)
Maxillofacial system	Oral fibrosis and stricture	0 (0%)	4 (20%)
	Reconstruction surgery	0 (0%)	1 (5%)
Plastic surgery	External chest wall burns	0 (0%)	1 (5%)
	Facial burns	0 (0%)	2 (10%)
Nutrition	Gastrostomy tube	0 (0%)	2 (10%)

Data are presented as n (%), with percentages calculated using the corresponding subgroup denominators: weak alkali, n=13; strong alkali, n=20. P values were calculated using Fisher's exact test based on the observed frequencies. Clinical complications are presented by study cohort. Endoscopic findings are illustrated separately, with percentages calculated from only those patients who underwent EGD (n=20), EGD: Esophagogastroduodenoscopy

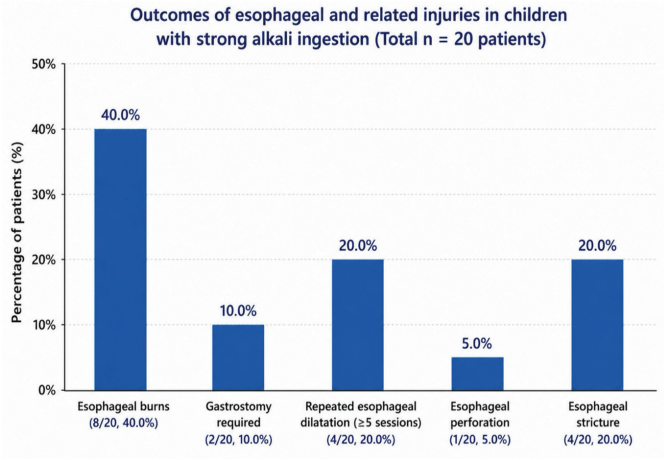


Figure 1. Endoscopic findings among children with strong alkaline ingestion. Percentages are calculated using only the 20 patients who underwent upper gastrointestinal endoscopy esophagogastroduodenoscopy as the denominator

incidence occurring at 2 years of age. The presentation of these children to the emergency department depends on the amount of alkaline substance ingested, the pH of the substance, and the time elapsed since ingestion. Household detergents and degreasers containing high concentrations of alkalis, with a pH as high as 11, are commonly found in lye-based products such as sodium hydroxide and potassium hydroxide drain cleaners, detergent powders, and granules (3,20). Ingestion of small amounts of weak alkalis or alkalis at low concentrations typically results in minor medical issues, such as mild vomiting and oral mucosal edema. Therefore, EGD is not always necessary in these patients, and close observation for a few hours is sufficient. Ingestion of a strong alkaline substance with a high pH is always associated with serious health consequences. Our current research highlighted the harmful effects of strong alkaline substances on multiple body organs. We observed that strong alkaline substances were ingested in 20 cases, while weak alkaline substances were ingested in 13 cases.

The pathophysiology of injuries caused by alkalis differs from that of acids. Alkalis can induce mucosal liquefactive necrosis, leading to submucosal destruction and deeper injuries that penetrate the muscularis propria. Alkalis become soluble upon forming soaps with fat, resulting in edematous tissue loosening and deep alkali diffusion into the tissues (3,20). This profound tissue damage results from the deep diffusion of alkalis and their edematous spread within the organs. Within 48 hours of ingestion, fibroblast proliferation and collagen synthesis commence, with peak weakening of the esophageal wall occurring between 1 and 3 weeks. The need for endoscopic evaluation should be determined based on clinical presentation and assessment by the treating pediatric gastroenterologist. Symptomatic children with suspected caustic gastrointestinal injury should undergo timely EGD, whereas routine endoscopy may not be necessary in asymptomatic children without oral lesions when adequate follow-up can be assured (21).

The most common serious complications of strong alkaline ingestion occur in the GIT. A solution of 30% sodium hydroxide can result in full-thickness injury within seconds (5). In a rat model of esophageal injury, the concentration of sodium hydroxide was found to be crucial in terms of damage capacity as 2-15% was sufficient to result in epithelial necrosis and adventitial necrosis while 34% result in damages to the lung, trachea and esophageal perforation within 10 minutes to two hours (22). Other complications documented in our study included oropharyngeal burns, esophageal burns, and gastric erosions.

Management should be initiated and performed only in hospitals where they have facilities for pediatric emergency services and pediatric gastroenterology services. The common practice in alkali ingestion whether strong or weak begins in the emergency room, with priorities to support airway, breathing, and circulation and to ensure hemodynamic stability. Treatment attempts in the acute stages of alkali ingestion by children are aimed at preventing the development of strictures (23). Children who have ingested weak alkalis may be monitored in the emergency room and provided with oral fluids as part of management. While endoscopy is not typically required for such patients, routine blood tests can be conducted. In our study, 13 individuals had a history of ingesting weak alkalis and presented with minimal vomiting or mild oral cutaneous edema. However, for those who have ingested strong alkalis, prompt management is crucial to safeguard the GIT and respiratory systems. Among our patients who ingested strong alkalis, 90% received intravenous dexamethasone as part of the initial treatment, although the use of steroid injections remains controversial. All patients were administered proton pump inhibitors and antibiotics. These treatments work together to mitigate further damage from acidic reflux, which could be detrimental in a compromised GIT, and to prevent secondary bacterial infections that may arise in necrotic tissue, potentially worsening the prognosis by leading to stricture formation.

The European Society of Gastrointestinal Endoscopy and European Society for Pediatric Gastroenterology Hepatology and Nutrition recommend EGD in children with suspected caustic ingestion who exhibit signs or symptoms, such as oral lesions, vomiting, drooling, dysphagia, hematemesis, dyspnea, or abdominal pain. In asymptomatic children without oral lesions, EGD may be withheld when adequate follow-up can be ensured (21). EGD is considered an essential procedure for grading cases and selecting the best management for patients with strong

alkaline ingestion. Previous research has raised doubts about the necessity of performing routine EGD in all suspected cases of caustic ingestion (24-26), arguing that significant injury is unlikely if symptoms do not manifest following alkaline ingestion. Crain et al. (27) examined the relationship between the presence or absence of three symptoms and signs (vomiting, drooling, and stridor) and the EGD findings, indicating that the presence of ≥2 signs or symptoms is a reliable predictor of esophageal injury. Although severe esophageal lesions have been observed in children who were initially asymptomatic, significant lesions at endoscopy have not been reported when symptoms are absent following ingestion of weak corrosives (28). Moreover, the need for EGD in children should be determined by a pediatric gastroenterologist. If EGD is required, it should be performed within 48 hours to prevent further organ damage. In our study, individuals who ingested strong alkalis underwent EGD within 48 hours of ingestion. We posit that children with a mild clinical presentation who ingested weakly alkaline substances and were assessed by a pediatric gastroenterologist may not require an EGD. Instead, close monitoring is deemed essential.

The majority of complications in our study affected the GIT and the respiratory tract. Other injuries involved the oral cavity, including the oropharynx, and the glottic and subglottic structures, with some lesions extending into the larynx. In cases where oropharynx and larynx are affected, flexible laryngoscopy is necessary to estimate the extent of injuries (29). In our study, respiratory involvement ranged from mild to severe, with two cases requiring admission to the pediatric ICU and intubation. Esophageal injury is considered the most critical and life-threatening complication of alkaline toxicity. These complications impose a socioeconomic burden on developing countries, were access to repeated interventions, such as esophageal dilatation, may be challenging. Patients may also experience significant nutritional depletion, requiring the use of a nasogastric tube (NGT) for long-term management. The tube serves as both a feeding route and a stent. In advanced esophageal injuries, stenting the damaged segment with an NGT may provide a route for enteral nutrition for a couple of weeks. In certain instances, such as those observed in our study, an esophageal stricture can develop. The incidence of stricture formation was higher in children who vomited after alkaline ingestion. These strictures are typically identified in the third week post-injury, and early dilatation may enhance the outcomes of esophageal injuries (28). Consequently, all patients with minimal esophageal damage should undergo follow-up endoscopy within three weeks of the initial detection of esophageal injury to exclude the possibility of stricture formation (16). In our study, esophageal stricture developed in 4 of the 20 children who underwent EGD (20%). Some researchers have reported that the rate of esophageal stricture formation can be as high as 40% or more in pediatric patients with moderate injuries following alkali ingestion (30). Regular esophageal dilatations should be performed to achieve an adequate esophageal caliber; however, the optimal frequency for esophageal balloon dilatation is not well-defined in the literature and is currently guided by the patient's symptoms.

Study Limitations

Only a few limitations have been acknowledged. As this was a retrospective medical-record review, some variables were not available for all patients. No statistical imputation was performed. The extent of missing data was reported where applicable. Furthermore, non-standardized clinical follow-up may have resulted in incomplete ascertainment of delayed complications.

Conclusion

This study describes the clinical and endoscopic characteristics of alkaline ingestion in children and highlights the spectrum of associated short- and long-term complications. Early endoscopic evaluation can assist in identifying and grading gastrointestinal injury and may support subsequent management decisions in appropriately selected patients. Given the retrospective design and absence of a comparison group based on endoscopy timing, the present study cannot determine whether early endoscopy reduces morbidity or mortality.

Ethics

Ethics Committee Approval: The study was conducted in accordance with the Declaration of Helsinki and was approved by the Biomedical Ethics Committee of King Abdulaziz University, Faculty of Medicine (approval no: 483-16. date: January 8, 2017).

Informed Consent: Informed consent was obtained from all parents before the procedure and the study.

Availability of Data and Material: The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflict of Interest: No conflict of interest was declared by the author.

Financial Disclosure: The author declared that this study received no financial support.

References

1. Cheng HT, Cheng CL, Lin CH, Tang JH, Chu YY, Liu NJ, et al. Caustic ingestion in adults: the role of endoscopic classification in predicting outcome. BMC Gastroenterol. 2008;8:31.

2. Brent J. Water-based solutions are the best decontaminating fluids for dermal corrosive exposures: a mini review. Clin Toxicol (Phila). 2013;51:731-6.
3. Arévalo-Silva C, Eliashar R, Wohlgelernter J, Elidan J, Gross M. Ingestion of caustic substances: a 15-year experience. Laryngoscope. 2006;116:1422-6.
4. Dehghani SM, Bahmanyar M, Javaherizadeh H. Caustic ingestion in children in South of Iran: a two-year single center study. Middle East J Dig Dis. 2018;10:31-4.
5. Lupa M, Magne J, Guarisco JL, Amedee R. Update on the diagnosis and treatment of caustic ingestion. Ochsner J. 2009;9:54-9.
6. Contini S, Swarray-Deen A, Scarpignato C. Oesophageal corrosive injuries in children: a forgotten social and health challenge in developing countries. Bull World Health Organ. 2009;87:950-4.
7. Gummin DD, Mowry JB, Beuhler MC, Spyker DA, Brooks DE, Dibert KW, et al. Annual report of the American Association of Poison Control Centers' National Poison Data System (NPDS): 37th Annual Report. Clin Toxicol (Phila). 2020;58:1360-541.
8. Hoffman RS, Howland MA, Kamerow HN, Goldfrank LR. Comparison of titratable acid/alkaline reserve and pH in potentially caustic household products. J Toxicol Clin Toxicol. 1989;27:241-6.
9. Karnak I, Tanyel FC, Büyükpamukçu N, Hiçsönmez A. Pulmonary effects of household bleach ingestion in children. Clin Pediatr (Phila). 1996;35:471-2.
10. Günel E, Cağlayan F, Cağlayan O, Akillioğlu I. Reactive oxygen radical levels in caustic esophageal burns. J Pediatr Surg. 1999;34:405-7.
11. Schrage NF, Langefeld S, Zschocke J, Kuckelkorn R, Redbrake C, Reim M. Eye burns: an emergency and continuing problem. Burns. 2000;26:689-99.
12. Chen RJ, O'Malley RN, Salzman M. Updates on the evaluation and management of caustic exposures. Emerg Med Clin North Am. 2022;40:343-64.
13. Shcherbaeva T, Muallem Kalmovich L, Shalem T, Broide E, Gavriel H, Pitaro J. Caustic ingestion in children: the otolaryngologist perspective. Pediatr Emerg Care. 2022;38:e1541-e4.
14. Zargar SA, Kochhar R, Mehta S, Mehta SK. The role of fiberoptic endoscopy in the management of corrosive ingestion and modified endoscopic classification of burns. Gastrointest Endosc. 1991;37:165-9.
15. Khan S, Orenstein SR. Ingestions. In: Kliegman RM, Stanton BF, Geme JW, Schor NF, Behrman RE, eds. Nelson Textbook of Pediatrics. 20th ed. Philadelphia: WB Saunders; 2016:1794-6.
16. Gharib B, Mohammadpour M, Yaghmaie B, Sharifzadeh M, Mehdizadeh M, Zamani F, et al. Caustic agent ingestion by a 1.5-year-old boy. Acta Med Iran. 2016;54:465-70.
17. Tiryaki T, Livanelioğlu Z, Atayurt H. Early bougienage for relief of stricture formation following caustic esophageal burns. Pediatr Surg Int. 2005;21:78-80.
18. Chirica M, Bonavina L, Kelly MD, Sarfati E, Cattan P. Caustic ingestion. Lancet. 2017;389(10083):2041-52.
19. Rafeey M, Ghojazadeh M, Mehdizadeh A, Hazrati H, Vahedi L. Intercontinental comparison of caustic ingestion in children. Korean J Pediatr. 2015;58:491-500.
20. Riffat F, Cheng A. Pediatric caustic ingestion: 50 consecutive cases and a review of the literature. Dis Esophagus. 2009;22:89-94.
21. Tringali A, Thomson M, Dumonceau JM, Tavares M, Tabbers MM, Furlano R, et al. Pediatric gastrointestinal endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) and European Society for Paediatric Gastroenterology Hepatology and Nutrition (ESPGHAN) Guideline Executive summary. Endoscopy. 2017;49:83-91.
22. Emoto Y, Yoshizawa K, Shikata N, Tsubura A, Nagasaki Y. Autopsy results of a case of ingestion of sodium hydroxide solution. J Toxicol Pathol. 2016;29:45-7.
23. Karaman İ, Koç O, Karaman A, Erdoğan D, Çavuşoğlu YH, Afşarlar ÇE, et al. Evaluation of 968 children with corrosive substance ingestion. Indian J Crit Care Med. 2015;19:714-8.
24. Thomson M, Tringali A, Dumonceau JM, Tavares M, Tabbers MM, Furlano R, et al. Paediatric gastrointestinal endoscopy: European Society for Paediatric Gastroenterology Hepatology and Nutrition and European Society of Gastrointestinal Endoscopy Guidelines. J Pediatr Gastroenterol Nutr. 2017;64:133-53.
25. Gorman RL, Khin-Maung-Gyi MT, Klein-Schwartz W, Oderda GM, Benson B, Litovitz T, et al. Initial symptoms as predictors of esophageal injury in alkaline corrosive ingestions. Am J Emerg Med. 1992;10:189-94.
26. Nuutinen M, Uhari M, Karvali T, Kouvalainen K. Consequences of caustic ingestions in children. Acta Paediatr. 1994;83:1200-5.
27. Crain EF, Gershel JC, Mezey AP. Caustic ingestions. Symptoms as predictors of esophageal injury. Am J Dis Child. 1984;138:863-5.
28. Boskovic A, Stankovic I. Predictability of gastroesophageal caustic injury from clinical findings: is endoscopy mandatory in children? Eur J Gastroenterol Hepatol. 2014;26:499-503.
29. Cowan D, Ho B, Sykes KJ, Wei JL. Pediatric oral burns: a ten-year review of patient characteristics, etiologies and treatment outcomes. Int J Pediatr Otorhinolaryngol. 2013;77:1325-8.
30. Elshabrawi M, A-Kader HH. Caustic ingestion in children. Expert Rev Gastroenterol Hepatol. 2011;5:637-45.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.85451
Eurasian J Emerg Med. 2026;25: 456-66

Performance of Consumer Multimodal Large Language Model Interfaces on a Selected Publication-derived Challenge Set of Initially Misdiagnosed Dermatologic Emergencies

Bengisu Meral Ketenci[1], Demet Acar[2]

[1]Konya City Hospital, Clinic of Dermatology, Konya, Türkiye
[2]Konya City Hospital, Clinic of Emergency Medicine, Konya, Türkiye

Abstract

Aim: To evaluate the diagnostic and safety-related response performance of consumer multimodal large language model interfaces on a selected publication-derived challenge set of initially misdiagnosed dermatologic emergencies.

Materials and Methods: This exploratory cross-sectional study evaluated 50 cases of initially misdiagnosed dermatologic emergencies, selected from published reports identified between May 1 and June 1, 2026. ChatGPT, Gemini, and Claude were tested on June 2, 2026, under text-only and text-plus-image conditions using a single standardized prompt. Outcomes were Top-1 and cumulative Top-4 accuracies, clinically appropriate emergency management, risk-triage failure, and harmful recommendations.

Results: Top-1 accuracy ranged from 32.0% to 50.0% with text-only input and from 36.0% to 58.0% with text-plus-image input; corresponding Top-4 ranges were 50.0%-66.0% and 56.0%-74.0%. Appropriate management ranged from 38.0% to 54.0% and from 44.0% to 64.0%. No within-platform input-mode comparisons remained significant after Holm correction. Between-platform differences remained significant for Top-1 accuracy under both input conditions and for Top-4 accuracy and management appropriateness under text-plus-image input. Post-hoc differences primarily reflected lower observed performance for Claude than for Gemini in this dataset. Appropriate management occurred in 100% of Top-1-correct responses compared with 9.9% of Top-1-incorrect responses, whereas 18.0% of Top-4-correct responses had inappropriate management.

Conclusion: Within this selected challenge set and a single standardized prompting condition, the platforms demonstrated moderate diagnostic performance and variable performance on safety-related responses. Correct Top-1 diagnosis closely corresponded to appropriate management, whereas Top-4 recognition did not ensure safe management. The findings represent a single-run technical benchmark and do not establish output reproducibility, clinical usefulness, real-world safety, or stable platform superiority.

Keywords: Dermatologic emergencies, large language models, artificial intelligence, diagnostic accuracy, patient safety, emergency management

Introduction

Dermatologic emergencies represent a diagnostically challenging subset of skin diseases in which delayed recognition or inappropriate initial management may lead to severe morbidity, permanent sequelae, or mortality (1). Although many dermatologic complaints are managed in outpatient settings, certain presentations require urgent evaluation and rapid intervention (2). Severe cutaneous adverse reactions, necrotizing soft tissue infections, disseminated viral eruptions, purpura fulminans, erythroderma, acute autoimmune blistering diseases, and airway-threatening angioedema are examples of conditions in which early clinical suspicion directly affects patient outcomes (1,2).

The diagnosis of dermatologic emergencies is often complicated by overlapping clinical morphology (1). Life-threatening diseases may initially resemble common benign conditions such as viral

Corresponding Author: Bengisu Meral Ketenci MD, Konya City Hospital, Clinic of Dermatology, Konya, Türkiye
E-mail: meralbengisu@gmail.com **ORCID ID:** orcid.org/0009-0007-5339-1681
Cite this article as: Ketenci BM, Acar D. Performance of consumer multimodal large language model interfaces on a selected publication-derived challenge set of initially misdiagnosed dermatologic emergencies. Eurasian J Emerg Med. 2026;25: 456-66

Received: 15.07.2026
Accepted: 22.09.2026
Published: 09.10.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

exanthems, urticaria, cellulitis, eczema, or uncomplicated drug eruptions (3). Conversely, some benign dermatoses may appear alarming at first presentation (2). This diagnostic overlap creates a risk of both delayed treatment and inappropriate triage (4). Therefore, in emergency dermatology, accurate diagnosis alone is not sufficient; clinicians must also recognize disease severity, identify red flags, recommend appropriate investigations, and determine whether urgent referral, hospitalization, or multidisciplinary consultation is required (1,5).

Large language models (LLMs) have been investigated as potential clinical decision-support technologies (6). With the development of multimodal systems, these models can process both textual clinical information and visual data, including clinical photographs (7). This capability is relevant to dermatology, where diagnostic reasoning depends on integrating lesion morphology, distribution, symptom history, medication exposure, systemic findings, and visual pattern recognition (8). Previous research has largely focused on diagnostic performance using clinical text, images, or combined text-image inputs (9,10). However, performance on retrospective benchmarks should be distinguished from reproducibility, clinical usefulness, and safety in real-world implementation.

Beyond diagnostic classification, dermatologic emergencies require recognition of disease severity, red flags, and the need for urgent escalation. A model may include the correct diagnosis in its differential list yet provide unsafe reassurance, omit hospitalization, or suggest inappropriate treatment (10). Therefore, safety-related response endpoints such as management appropriateness, risk-triage failures, and harmful recommendations can complement diagnostic accuracy when benchmarking model outputs. These response-based endpoints do not, however, constitute evidence of clinical safety in actual implementation.

This study aimed to benchmark the diagnostic and safety-related response performance of three consumer multimodal LLM interfaces using a selected, publication-derived challenge set comprising 50 initially misdiagnosed dermatologic emergency cases. Each case was assessed under paired text-only and text-plus-image conditions using a single standardized prompt. The study directly evaluated performance on this selected case set but did not evaluate output reproducibility, clinical effectiveness as a decision-support tool, or safety in real-world implementation.

Materials and Methods

Study Design

This exploratory cross-sectional comparative study benchmarked the diagnostic and safety-related response performance of three consumer multimodal LLM interfaces using a selected publication-derived challenge set comprising initially misdiagnosed dermatologic emergencies. The dataset comprised 50 cases identified from published case reports and case series using the literature search and eligibility assessment process described below. All included cases had been initially misdiagnosed, and the final diagnoses had been confirmed by histopathology, microbiology, laboratory testing, imaging, genetic analysis, or accepted clinical criteria.

The literature search and dataset construction were conducted between May 1 and June 1, 2026, and model evaluations were performed on June 2, 2026. Each case was assessed under two separately administered, paired input conditions: text-only and text-plus-image. Outcomes included Top-1 accuracy, cumulative Top-4 accuracy, clinically appropriate emergency management, risk-triage failure, and harmful recommendations.

Case Selection and Source-publication Provenance

A structured literature search was conducted in PubMed/MEDLINE, Embase, and Google Scholar between May 1 and June 1, 2026, without a publication-date restriction and was limited to English-language publications. Search terms included the following: dermatologic emergencies, specific urgent conditions, diagnostic errors, case reports, and case series. The complete search strategy is provided in Supplementary Table S1.

Eligible publications were required to report at least one case with an explicitly documented initial misdiagnosis, a supported final diagnosis, sufficient clinical information, and at least one suitable clinical image. An emergency or urgent presentation was defined as an acute or rapidly progressive cutaneous or mucocutaneous condition in which delayed management could result in death, airway or organ compromise, severe deterioration, irreversible damage, or the need for urgent hospital-based intervention. Publications not meeting these criteria were excluded.

The complete selection process and the numerical reasons for exclusion are presented in Figure 1. The initial search was performed at the publication level and included both case reports and case series as potentially eligible publication types. Full-text eligibility assessment was also performed at the publication level. All 50 publications that ultimately satisfied the eligibility criteria were single-patient case reports; therefore, each included publication contributed exactly one case to the final dataset (n=50 publications; n=50 cases). Case-level source and provenance information is provided in Supplementary Table S2. Eligibility was independently assessed by a board-certified dermatologist and an emergency medicine specialist (Cohen's κ =0.83), with disagreements resolved by consensus.

Disease Categories

Cases were classified into six mutually exclusive categories: severe cutaneous adverse reactions (n=12), infectious emergencies (n=11), vascular or vasculitic emergencies (n=9), autoimmune blistering emergencies (n=8), urticaria/angioedema disorders (n=6), and other dermatologic emergencies (n=4).

Artificial Intelligence Models

The default free consumer web versions of ChatGPT, Gemini, and Claude were evaluated on June 2, 2026, from Türkiye via their official web interfaces. Testing was conducted in private (incognito) browser sessions without selecting a named model or using paid subscription features. Because the consumer interfaces did not provide verifiable backend metadata, the exact backend model versions were unavailable and the systems were identified only by their platform names.

API access, plugins, fine-tuning, retrieval-augmented generation, automated web search, externally connected tools, and custom system instructions were not used. No explicit rate-limit or fallback-model notification was encountered during testing.

A total of 300 model-case-input-condition evaluations were performed. Each evaluation was conducted on a separate, newly opened page and in an independent conversation to prevent information carryover between cases or input conditions. Text-only and text-plus-image evaluations of the same case were performed separately; each condition was queried once.

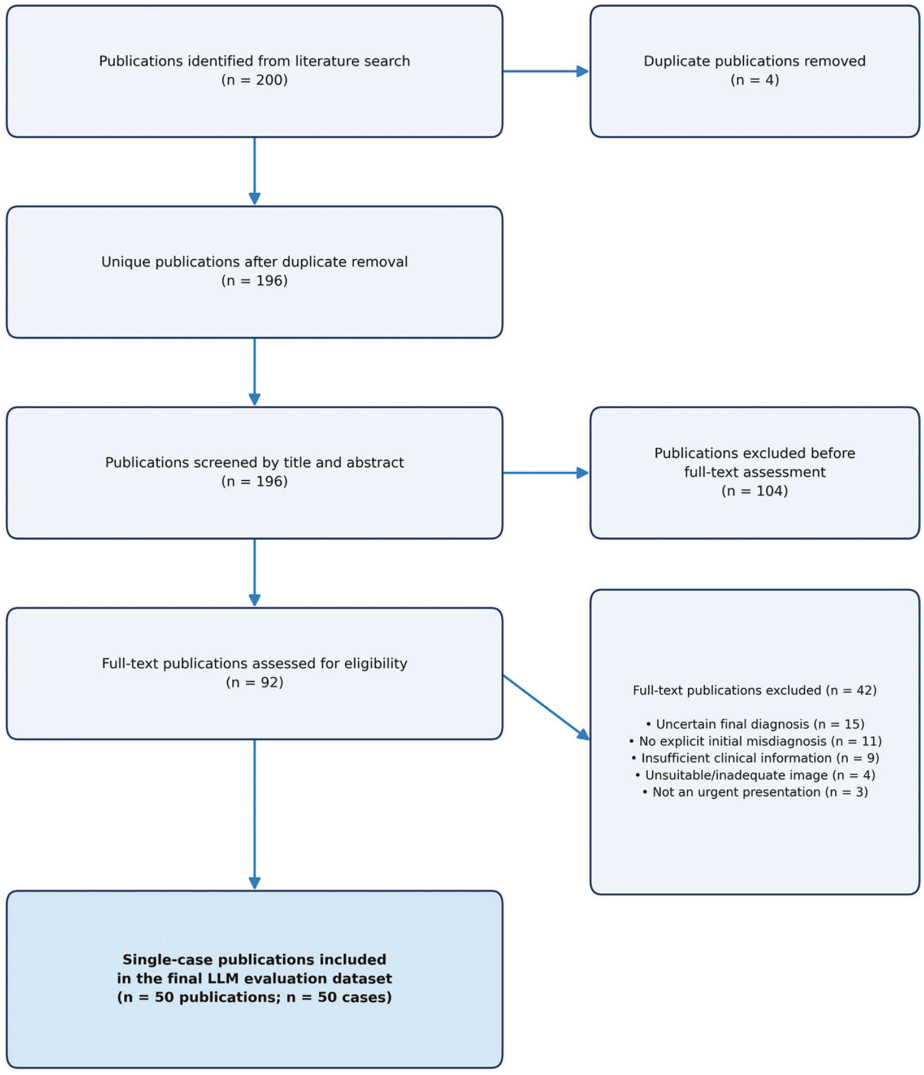


Figure 1. Flow diagram of the literature search and publication selection process. Publications were screened at the title/abstract and full-text stages. All 50 publications included in the final LLM evaluation dataset were single-patient case reports; therefore, the final dataset comprised 50 publications corresponding to 50 cases

LLM: Large language model

Prompting Protocol

In the text-only condition, each model received a clinical vignette prepared using the standardized construction framework and containing demographic, historical, systemic, and physical examination findings.

In the text-plus-image condition, the same prepared vignette and the clinical image from the original publication were provided in a separate conversation.

The standardized prompt was:

"You are an expert emergency physician. Analyze the following patient presenting to the emergency department with a dermatologic complaint using the provided clinical history (and image). Provide the most likely diagnosis, three differential diagnoses ranked in order of probability, and the recommended clinical approach. Use medical terminology."

Clinical Vignette Construction

Clinical vignettes were manually prepared from the source publications using using a standardized construction framework. The investigators aimed to include information considered to be available at the initial clinical presentation, including demographic characteristics, presenting symptoms and their duration, relevant medical and medication history, systemic symptoms, vital signs, and physical and dermatologic examination findings. The standardized construction framework specified that final diagnoses, diagnostic-confirmation findings, and post-diagnostic laboratory, histopathological, microbiological, imaging, therapeutic response, and follow-up information were not included in the vignettes. The initial incorrect diagnosis was not intentionally presented as a diagnostic label. The source material was paraphrased and summarized, and publication titles, citations, figure captions, annotations, and direct source-identifying statements were not included in the prompts. A consistent set of clinical fields was sought, although no fixed word-count requirement was applied, because the completeness of the published reports varied.

Response Evaluation

Responses were independently assessed by a dermatologist and an emergency medicine specialist, who were blinded to each other's scores during the initial evaluation phase. After completion of independent scoring, all discordant ratings were jointly reviewed and resolved through structured discussion and consensus. The final consensus ratings, rather than either assessor's individual ratings, were used in the main analyses. No third assessor was involved, and neither assessor's rating was used as a default decision.

The final diagnosis reported in the source publication was used as the diagnostic reference standard. Emergency management was evaluated according to current literature, clinical guidelines, and established practice principles.

Diagnostic Accuracy

Top-1 accuracy was defined as the first-ranked diagnosis being identical to or clinically equivalent to the reference diagnosis.

Top-4 accuracy was defined as a cumulative measure indicating that the reference diagnosis was included either as the most likely diagnosis or as one of the three ranked differential diagnoses. Therefore, a correct Top-1 response was also considered correct in the Top-4.

Patient-safety Outcomes

Risk-triage failure was defined as the failure to recognize or communicate urgency, including underestimating severity, omitting urgent referral or hospitalization, or failing to identify key red flags.

A harmful recommendation was defined as an active suggestion that was incorrect, contraindicated, falsely reassuring, potentially harmful, or capable of delaying urgent care.

These outcomes were not mutually exclusive. Assessment criteria were determined on a disease-specific basis. Relevant Turkish national treatment guidelines were prioritized when available. In their absence, widely accepted international clinical practice guidelines or consensus statements were used. When no applicable guideline or consensus statement was available, the relevant literature and the clinical judgment and experience of the dermatologist and the emergency medicine specialist were considered. The two assessors scored the responses independently, and disagreements were resolved by consensus.

Clinically Appropriate Emergency Management

Clinically appropriate emergency management was evaluated using a two-step approach. For brevity, this outcome is referred to as "management appropriateness" throughout the manuscript. Responses containing a risk-triage failure or a harmful recommendation were classified as inappropriate. Responses without these safety violations were assessed for adequate recognition of urgency, referral or hospitalization, specialist consultation, initial treatment, diagnostic work-up, and safety warnings.

Ethical Considerations

The literature search and dataset construction were conducted between May 1 and June 1, 2026, and model evaluations were performed on June 2, 2026. The study used only anonymized

clinical information and images obtained from published case reports and case series. No patient intervention was performed, no new clinical data were collected, no identifiable personal data were used, and no biological material sampling was performed.

Because the study was based on a secondary analysis of previously published anonymized data and did not involve the collection of new patient data or the inclusion of identifiable personal information, it was considered a study that did not require ethics committee approval or patient informed consent.

All case information was reviewed prior to analysis to remove elements that could identify patients. The study was conducted in accordance with the 2013 revision of the Declaration of Helsinki, ICMJE recommendations, and COPE guidelines. Clinical images obtained from original publications were used with appropriate citation of the relevant sources and in accordance with the licensing and copyright conditions of the publications.

Before submission to the model interfaces, clinical images were reviewed and cropped to include only the dermatologic lesion where possible. Faces, tattoos, unique markings, personal identifiers, publication captions, labels, arrows, and watermarks were removed when present. The images were processed using default consumer web interfaces in private (incognito) browser sessions.

Statistical Analysis

Top-1 accuracy and clinically appropriate emergency management were designated as primary outcomes, whereas cumulative Top-4 accuracy, risk-triage failure, and harmful recommendation were designated as secondary outcomes. This reporting hierarchy was not preregistered.

All 50 eligible cases identified during dataset construction were included; therefore, no formal a priori power calculation was performed. With 50 cases, a proportion of 50% corresponds to a Wilson 95% confidence interval (CI) of 36.6%-63.4%, indicating a maximum precision of approximately ±13.4 percentage points.

Outcomes were summarized as n/N (%) with Wilson 95% CIs. Within each platform, text-only and text-plus-image conditions were compared using two-sided exact McNemar tests. Paired risk differences, calculated as text-plus-image minus text-only, were reported with 95% CIs calculated using the Newcombe hybrid-score method. Comparisons among the three platforms within each input condition were performed using Cochran's Q test. Paired between-platform risk differences, with 95% CIs, were additionally reported. When an omnibus comparison was significant, post-hoc pairwise comparisons were conducted using exact McNemar tests.

Multiplicity was addressed using the Holm procedure across the 25 principal comparisons, comprising 15 within-platform McNemar tests and 10 omnibus Cochran Q tests. Post-hoc pairwise p-values were Bonferroni-adjusted within each family of three comparisons. CIs were not adjusted for multiplicity and were interpreted as measures of estimation precision. Statistical significance was interpreted according to exact and Holm-adjusted p-values, rather than unadjusted CIs. Non-significant findings were not interpreted as evidence of equivalence or absence of an effect.

Category-specific analyses were descriptive because of the small subgroup sizes and were reported as n/N with whole-number percentages. Diagnostic-management concordance was evaluated using response-level cross-tabulations. Conditional outcome rates and risk differences with 95% CIs were calculated according to Top-1 correctness, with an exploratory analysis based on Top-4 accuracy. Inter-assessor agreement was calculated using the independent pre-consensus ratings. Final outcome counts and all inferential analyses were based on the final consensus ratings. The relationship between the independent ratings and the final consensus ratings is provided at the case level in Supplementary Data S1. Analyses were performed using IBM SPSS Statistics, version 29.0, and independently verified using the case-level dataset.

Results

A total of 50 initially misdiagnosed dermatologic emergency cases were evaluated across three consumer multimodal LLM platforms under text-only and text-plus-image conditions, corresponding to 300 response-level evaluations. Overall outcomes, with Wilson 95% CIs, are presented in Table 1. Paired input-mode comparisons (Table 2), omnibus between-platform comparisons (Table 3), and post-hoc pairwise comparisons (Table 4) are presented. The case-selection process is shown in Figure 1.

Overall Performance

Top-1 accuracy ranged from 16/50 (32.0%; 95% CI 20.8%-45.8%) to 25/50 (50.0%; 95% CI 36.6%-63.4%) under text-only input and from 18/50 (36.0%; 95% CI 24.1%-49.9%) to 29/50 (58.0%; 95% CI 44.2%-70.6%) under text-plus-image input. Top-4 accuracy ranged from 25/50 (50.0%; 95% CI 36.6%-63.4%) to 33/50 (66.0%; 95% CI 52.2%-77.6%) and from 28/50 (56.0%; 95% CI 42.3%-68.8%) to 37/50 (74.0%; 95% CI 60.4%-84.1%), respectively.

Clinically appropriate emergency management ranged from 19/50 (38.0%; 95% CI 25.9%-51.8%) to 27/50 (54.0%; 95% CI 40.4%-67.0%) under text-only input, and from 22/50 (44.0%; 95% CI 31.2%-57.7%) to 32/50 (64.0%; 95% CI 50.1%-75.9%) under text-plus-image input (Table 1).

Table 1. Overall outcomes by platform and input condition

Platform	Input	Top-1 accuracy	Top-4 accuracy	Appropriate management	Risk-triage failure	Harmful recommendation
ChatGPT	Text-only	23/50 (46.0%; 33.0-59.6)	31/50 (62.0%; 48.2-74.1)	25/50 (50.0%; 36.6-63.4)	13/50 (26.0%; 15.9-39.6)	5/50 (10.0%; 4.3-21.4)
ChatGPT	Text-plus-image	28/50 (56.0%; 42.3-68.8)	35/50 (70.0%; 56.2-80.9)	30/50 (60.0%; 46.2-72.4)	10/50 (20.0%; 11.2-33.0)	3/50 (6.0%; 2.1-16.2)
Gemini	Text-only	25/50 (50.0%; 36.6-63.4)	33/50 (66.0%; 52.2-77.6)	27/50 (54.0%; 40.4-67.0)	12/50 (24.0%; 14.3-37.4)	4/50 (8.0%; 3.2-18.8)
Gemini	Text-plus-image	29/50 (58.0%; 44.2-70.6)	37/50 (74.0%; 60.4-84.1)	32/50 (64.0%; 50.1-75.9)	8/50 (16.0%; 8.3-28.5)	2/50 (4.0%; 1.1-13.5)
Claude	Text-only	16/50 (32.0%; 20.8-45.8)	25/50 (50.0%; 36.6-63.4)	19/50 (38.0%; 25.9-51.8)	18/50 (36.0%; 24.1-49.9)	8/50 (16.0%; 8.3-28.5)
Claude	Text-plus-image	18/50 (36.0%; 24.1-49.9)	28/50 (56.0%; 42.3-68.8)	22/50 (44.0%; 31.2-57.7)	15/50 (30.0%; 19.1-43.8)	6/50 (12.0%; 5.6-23.8)

Values are presented as n/N (%; Wilson 95% confidence interval)

Table 2. Paired text-only versus text-plus-image comparisons

Platform	Outcome	Text	Image	T=1/I=0	T=0/I=1	RD (95% CI), pp	Exact p	Holm p
ChatGPT	Top-1	23	28	0	5	+10 (+1.8 to +17.8)	0.063	1.000
ChatGPT	Top-4	31	35	1	5	+8 (−1.3 to +17.2)	0.219	1.000
ChatGPT	Management	25	30	1	6	+10 (+0.1 to +19.4)	0.125	1.000
ChatGPT	Risk-triage failure	13	10	7	4	−6 (−19.0 to +7.0)	0.549	1.000
ChatGPT	Harmful recommendation	5	3	5	3	−4 (−16.3 to +8.1)	0.727	1.000
Gemini	Top-1	25	29	0	4	+8 (+0.6 to +15.1)	0.125	1.000
Gemini	Top-4	33	37	0	4	+8 (+0.3 to +15.9)	0.125	1.000
Gemini	Management	27	32	0	5	+10 (+1.7 to +18.0)	0.063	1.000
Gemini	Risk-triage failure	12	8	5	1	−8 (−18.3 to +1.7)	0.219	1.000
Gemini	Harmful recommendation	4	2	3	1	−4 (−14.3 to +5.2)	0.625	1.000
Claude	Top-1	16	18	2	4	+4 (−5.5 to +13.4)	0.688	1.000
Claude	Top-4	25	28	1	4	+6 (−2.4 to +14.2)	0.375	1.000
Claude	Management	19	22	2	5	+6 (−4.1 to +15.8)	0.453	1.000
Claude	Risk-triage failure	18	15	8	5	−6 (−19.6 to +7.9)	0.581	1.000
Claude	Harmful recommendation	8	6	6	4	−4 (−17.0 to +8.9)	0.754	1.000

Risk differences were calculated as the difference between the text-plus-image and text-only conditions. Unadjusted Newcombe hybrid-score 95% confidence intervals were used. Exact p values are derived from two-sided exact McNemar tests and may not be inferentially equivalent to Newcombe confidence intervals. Statistical significance was interpreted using exact and Holm-adjusted p values rather than unadjusted confidence intervals. Holm p values refer to the 25 principal comparisons

Paired Input-mode Comparisons

For all three platforms, Top-1 accuracy, Top-4 accuracy, and clinically appropriate management were numerically higher with text-plus-image input, while risk-triage failure rates and harmful recommendation rates were lower. The paired risk differences ranged from +4 to +10 percentage points for Top-1 accuracy, +6 to +8 percentage points for Top-4 accuracy, and +6 to +10 percentage points for management appropriateness. However, none of the 15 paired input-mode comparisons was statistically significant after Holm correction (all adjusted p=1.000; Table 2). These non-significant findings were not interpreted as demonstrating equivalence between input conditions.

Between-platform Comparisons

Significant between-platform differences remained after Holm correction for Top-1 accuracy under text-only input (Q=13.400, adjusted p=0.028) and text-plus-image input (Q=15.857, adjusted p=0.009). Under text-plus-image input, significant differences were also observed for Top-4 accuracy (Q=13.400, adjusted p=0.028) and for clinically appropriate management (Q=14.000, adjusted p=0.022). The remaining omnibus comparisons were not statistically significant after multiplicity correction (Table 3).

In Bonferroni-adjusted post-hoc comparisons, Gemini had higher Top-1 accuracy than Claude under text-only input, with a paired

Table 3. Omnibus between-platform comparisons within each input condition

Input	Outcome	ChatGPT	Gemini	Claude	Cochran Q	Raw p	Holm p
Text-only	Top-1	23/50	25/50	16/50	13.400	0.001	0.028
Text-only	Top-4	31/50	33/50	25/50	8.667	0.013	0.262
Text-only	Management	25/50	27/50	19/50	9.455	0.009	0.186
Text-only	Risk-triage failure	13/50	12/50	18/50	4.133	0.127	1.000
Text-only	Harmful recommendation	5/50	4/50	8/50	2.600	0.273	1.000
Text-plus-image	Top-1	28/50	29/50	18/50	15.857	<0.001	0.009
Text-plus-image	Top-4	35/50	37/50	28/50	13.400	0.001	0.028
Text-plus-image	Management	30/50	32/50	22/50	14.000	<0.001	0.022
Text-plus-image	Risk-triage failure	10/50	8/50	15/50	5.571	0.062	1.000
Text-plus-image	Harmful recommendation	3/50	2/50	6/50	2.889	0.236	1.000

Cochran Q tests accounted for the fact that the same 50 cases were evaluated by all three platforms. The Holm correction was applied across the 25 principal comparisons

Table 4. Post-hoc pairwise comparisons for Holm-significant omnibus outcome families

Input/outcome	Comparison	RD, pp	95% CI, pp	Raw p	Bonferroni p
Text-only Top-1	ChatGPT-Gemini	−4	−9.2 to +1.3	0.500	1.000
Text-only Top-1	ChatGPT-Claude	+14	+2.9 to +24.5	0.039	0.117
Text-only Top-1	Gemini-Claude	+18	+7.2 to +27.9	0.004	0.012
Text-plus-image Top-1	ChatGPT-Gemini	−2	−10.5 to +6.5	1.000	1.000
Text-plus-image Top-1	ChatGPT-Claude	+20	+7.4 to +31.3	0.006	0.019
Text-plus-image Top-1	Gemini-Claude	+22	+10.3 to +32.4	0.001	0.003
Text-plus-image Top-4	ChatGPT-Gemini	−4	−12.2 to +4.0	0.625	1.000
Text-plus-image Top-4	ChatGPT-Claude	+14	+4.3 to +23.3	0.016	0.047
Text-plus-image Top-4	Gemini-Claude	+18	+7.2 to +28.2	0.004	0.012
Text-plus-image management	ChatGPT-Gemini	−4	−11.6 to +3.7	0.625	1.000
Text-plus-image management	ChatGPT-Claude	+16	+4.4 to +26.7	0.021	0.064
Text-plus-image management	Gemini-Claude	+20	+8.7 to +30.2	0.002	0.006

Risk differences were calculated as the values from the first platform minus those from the second platform. Confidence intervals are unadjusted. Pairwise p values were Bonferroni-adjusted within each three-comparison family, CI: Confidence interval

risk difference of 18 percentage points (95% CI, 7.2-27.9; adjusted p=0.012). Under text-plus-image input, ChatGPT and Gemini had higher Top-1 accuracy than Claude by 20 percentage points (95% CI 7.4–31.3; adjusted p=0.019) and 22 percentage points (95% CI 10.3-32.4; adjusted p=0.003), respectively. For text-plus-image Top-4 accuracy, ChatGPT and Gemini exceeded Claude by 14 percentage points (95% CI 4.3-23.3; adjusted p=0.047) and 18 percentage points (95% CI 7.2-28.2; adjusted p=0.012), respectively. Regarding management appropriateness, Gemini exceeded Claude by 20 percentage points (95% CI 8.7-30.2; adjusted p=0.006). No post hoc comparisons between ChatGPT and Gemini were statistically significant (Table 4).

Category-specific Management

Category-specific management findings were descriptive. When given text-plus-image input, ChatGPT and Gemini each provided appropriate management in 9/12 (75%) severe cutaneous adverse reaction cases. The appropriateness of management in vascular or vasculitic emergencies ranged from 3/9 (33%) to 5/9 (56%), depending on the platform and input condition. Because subgroup denominators ranged from four to twelve, these findings are considered exploratory and are presented as n/N with whole-number percentages in Figure 2.

Patient-safety Outcomes

Risk-triage failure ranged from 12/50 (24.0%; 95% CI 14.3%-37.4%) to 18/50 (36.0%; 95% CI 24.1%-49.9%) under text-only input and from 8/50 (16.0%; 95% CI 8.3%-28.5%) to 15/50 (30.0%; 95% CI 19.1%-43.8%) under text-plus-image input. Harmful recommendation rates ranged from 4/50 (8.0%; 95% CI 3.2%-18.8%) to 8/50 (16.0%; 95% CI 8.3%-28.5%) and from 2/50 (4.0%; 95% CI 1.1%-13.5%) to 6/50 (12.0%; 95% CI 5.6%-23.8%),

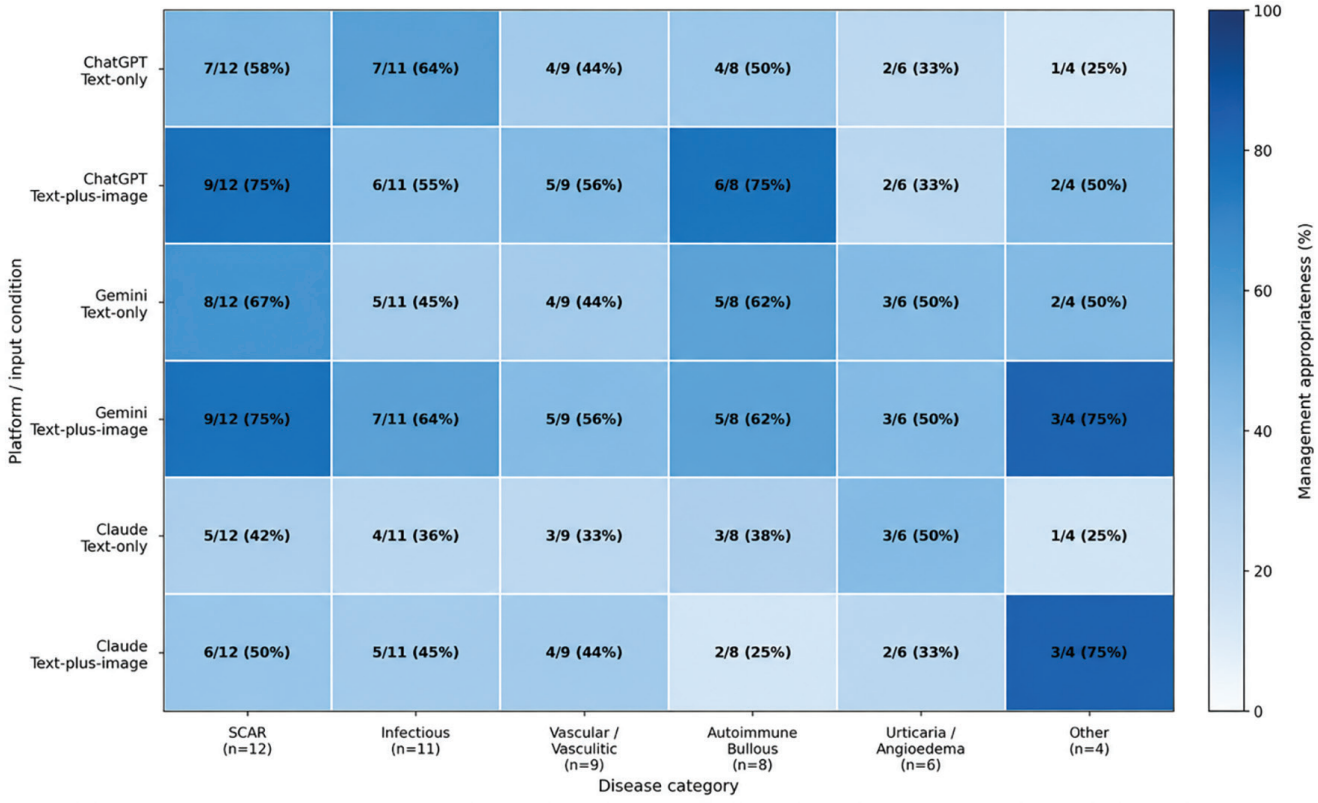


Figure 2. Clinically appropriate emergency management by disease category and input condition. Heat map showing n/N (whole-number percentage) for each platform, input condition, and disease category. A colour-blind-accessible, sequential blue palette is used, with darker shades indicating greater appropriateness of management

respectively. Neither outcome differed significantly between input conditions nor among platforms after the Holm correction (Tables 1-3).

Diagnostic-management Concordance

Diagnostic-management concordance was evaluated using a descriptive, response-level cross-tabulation across all 300 generated responses. For Top-1 accuracy, 139 responses were Top-1 correct with appropriate management, 0 were Top-1 correct with inappropriate management, 16 were Top-1 incorrect with appropriate management, and 145 were Top-1 incorrect with inappropriate management. Risk-triage failures and harmful recommendations occurred exclusively in Top-1 incorrect responses. For cumulative Top-4 accuracy, 155 responses were Top-4 correct with appropriate management; 34 were Top-4 correct with inappropriate management; 0 were Top-4 incorrect with appropriate management; and 111 were Top-4 incorrect with inappropriate management. These response-level descriptive findings indicate that a correct Top-1 diagnosis closely corresponds with appropriate management, whereas inclusion of the reference diagnosis within the Top-4 does not ensure safe management. Detailed response-level cross-tabulations and effect estimates are provided in the supplementary material.

Discussion

This single-run technical benchmark evaluated diagnostic and safety-related response performance across three three consumer multimodal LLM platforms using a selected publication-derived challenge set. Three findings emerged. First, text-plus-image input produced numerically more favorable outcomes, but none of the paired input-mode comparisons remained significant after multiplicity correction. Second, several between-platform differences remained significant, primarily reflecting lower observed Top-1, Top-4, and management performance for Claude than for Gemini in this dataset. Third, Top-1 diagnostic accuracy closely corresponded with appropriate management, whereas inclusion of the correct diagnosis among the Top-4 did not ensure safe management.

The response-level analysis clarified the relationship between diagnosis and management. All 139 Top-1-correct responses were accompanied by appropriate management, whereas only 16 of 161 Top-1-incorrect responses were. Risk-triage failures and harmful recommendations occurred only when the Top-1 diagnosis was incorrect. Nevertheless, 34/189 of the Top-4-correct responses had inappropriate management. The findings,

therefore, do not support a general dissociation between Top-1 diagnostic correctness and safe management; rather, they indicate that lower-ranked recognition of the correct diagnosis does not necessarily translate into appropriate clinical action.

Clinical images were associated with numerical improvements of 4-10 percentage points in Top-1 accuracy, 6-8 percentage points in Top-4 accuracy, and 6-10 percentage points in management appropriateness. Rates of risk-triage failure and harmful recommendations also declined numerically. However, none of these paired changes remained significant after Holm correction. Image resolution, illumination, lesion framing, skin tone, and the ability of a single photograph to represent lesion distribution may influence multimodal performance (7-9). Because image acquisition was not standardized, the results should be interpreted as reflecting performance under heterogeneous, publication-derived image conditions, not as evidence that image input is generally superior.

After multiplicity correction, between-platform differences remained for Top-1 accuracy under both input conditions, and for Top-4 accuracy and management appropriateness under the text-plus-image input. Post-hoc findings primarily reflected lower observed performance for Claude than for Gemini; ChatGPT also exceeded Claude in text-plus-image Top-1 and Top-4 accuracy. No significant post-hoc difference was observed between ChatGPT and Gemini. These comparisons are sample- and date-specific. They should not be interpreted as a stable ranking of the platforms because the exact backend versions were unavailable, each condition was queried only once, and consumer interfaces change over time.

Safety-related response outcomes provided information beyond the diagnostic rank. Harmful recommendations were less frequent than risk-triage failures, but represented active advice that could delay appropriate care. This distinction is relevant in time-sensitive conditions such as necrotizing soft tissue infection and severe cutaneous adverse reactions (4,11). Similar work in other acute-care and patient-facing LLM evaluations has separated diagnostic correctness from urgency, red-flag recognition, management, and risk of harm (12-14). Nevertheless, the present rates reflect only the content of generated responses and should not be interpreted as observed patient harm or real-world safety estimates.

Category-specific findings were heterogeneous and explicitly descriptive. With only 4-12 cases per category, a single response changed the percentage by approximately 8-25 percentage points. The comparatively favorable management outcomes for severe cutaneous adverse reactions and less favorable outcomes in some vascular or vasculitic cases may reflect differences in the available clinical history, morphology, and systemic context, but the data cannot establish category-level superiority or causal explanations. Comparable dermatology-focused evaluations have likewise reported variable diagnostic performance across LLM platforms and evaluation settings (15-17).

The study's inferential scope must be separated into four domains. First, it directly assesses performance on a deliberately selected publication-derived challenge set. Second, because only one generation was obtained per condition, it does not assess output reproducibility or within-platform variability. Third, without a contemporaneous clinician comparator, it cannot determine the clinical usefulness as a decision-support tool, or whether use of the platforms improves clinician performance. Fourth, simulated responses to published cases cannot establish safety in real-world implementation. The safety-oriented endpoints, therefore, represent only a structured assessment of potentially unsafe content within the generated responses.

Automation bias and anchoring remain relevant considerations for future implementation research (18,19), but they were not directly evaluated here. Similarly, evidence from other diagnostic domains and human-AI studies cannot be used to infer benefit in this dataset (20,21). Future studies should use prospectively collected and representative cases, identifiable model versions, repeated generations, neutral and prespecified prompts, case-specific management standards, and contemporaneous clinician comparator groups. These design elements are necessary before conclusions about reproducibility, clinical benefit, or implementation safety can be made (22).

A strength of the study was the use of confirmed cases that posed documented diagnostic difficulties in their source reports and the separate evaluation of diagnostic and safety-related response endpoints. The same testing structure was applied across the three platforms, and case eligibility and response scoring involved both dermatology and emergency medicine expertise. These features support internal comparison within the challenge set but do not remove the limitations associated with retrospective case selection, single-run testing, and expert adjudication.

Study Limitations

This study employed a small, publication-derived challenge set, deliberately enriched with unusual and initially misdiagnosed dermatologic emergencies. It was neither representative nor consecutive, and therefore could not be used to estimate routine emergency department diagnostic accuracy. Subgroup sizes were small, and category-specific findings should be considered descriptive and hypothesis-generating.

Each platform-case-condition combination was queried once using a single standardized prompt on one testing date. The study therefore does not measure within-platform output variability, reproducibility across repeated generations, or longitudinal stability. The exact backend versions could not be verified through the consumer interfaces. Although the prompt did not identify the cases as dermatologic emergencies, the emergency-department setting and the role of expert emergency physicians may have encouraged urgency-oriented responses and may have limited comparisons with unprompted clinical use. In addition, the exact original vignette texts, complete verbatim model responses, query-specific timestamps, and complete query-level access metadata were not systematically archived. Although the retained case-level scoring dataset permits verification of the quantitative analyses, the original model responses cannot be independently re-adjudicated, and prompt-level auditability is limited.

All cases and images were derived from published reports. Training-data contamination, memorization, and recognition of distinctive clinical details or images cannot be excluded, even though web search functionality was not activated and source-identifying wording and image annotations were excluded when possible. Image acquisition and image quality were not standardized, and a publication-date sensitivity analysis could not be conducted because back-end knowledge cut-offs were unavailable. In addition, potential overlap with a previously published pilot study that used initially misdiagnosed dermatologic cases could not be definitively assessed because the pilot study did not provide a complete case-level source list or sufficient case identifiers. No vignettes, images, model outputs, scoring data, or statistical analyses were knowingly reused from that study; however, unrecognized overlap at the level of published source cases cannot be excluded.

Management appropriateness, risk-triage failure, and harmful recommendations were based on expert judgment. Although two specialists independently evaluated responses, and any disagreements were adjudicated by consensus, a formal case-specific reference-management checklist was not prospectively archived. Subjective outcome adjudication, therefore, remains a limitation.

No contemporaneous human-expert comparator was included. Initial misdiagnosis in the source reports was used only as a case-selection characteristic and did not represent a clinician–platform comparison. Accordingly, the study should be interpreted strictly as a technical benchmark of consumer interfaces using the selected dataset and as a partial analysis of safety-related response content. It does not establish reproducibility, clinical usefulness, clinician-support benefits, or safety in real-world implementation.

Conclusion

In this single-run technical benchmark of a selected publication-derived challenge set, the consumer multimodal LLM platforms demonstrated moderate diagnostic performance and incomplete safety-related performance in their responses. Text-plus-image input produced numerically more favorable outcomes; however, no pairwise input-mode comparison remained statistically significant after correction for multiple comparisons. Several between-platform differences were observed within this dataset, primarily reflecting lower performance for Claude than for Gemini; these findings do not establish stable platform superiority.

A correct Top-1 diagnosis closely corresponded with appropriate management, whereas inclusion of the correct diagnosis within the Top-4 did not ensure safe management. The findings apply only to the tested cases, consumer interfaces, testing date, single generation, and standardized prompting condition. They do not establish output reproducibility, clinical usefulness as decision support, or safety in real-world practice. Future studies should use representative prospective cases, identifiable model versions, repeated generations, predefined management standards, and contemporaneous clinician comparator groups.

Ethics

Ethics Committee Approval: The literature search and dataset construction were conducted between May 1 and June 1, 2026, and model evaluations were performed on June 2, 2026. The study used only anonymized clinical information and images obtained from published case reports and case series.

Informed Consent: No patient intervention was performed, no new clinical data were collected, no identifiable personal data were used, and no biological material sampling was performed.

Footnotes

Authorship Contributions

Concept: B.M.K., D.A., Design: B.M.K., D.A., Data Collection or Processing: B.M.K., D.A., Analysis or Interpretation: B.M.K., D.A., Literature Search: B.M.K., D.A., Writing: B.M.K.

Conflict of Interest: Demet Acar (Editor-in-Chief) is a member of the Editorial Board of the Eurasian Journal of Emergency Medicine. However, she was not involved in the editorial evaluation or decision-making process for this manuscript. The manuscript was evaluated by editors from different institutions. The other author declared no conflicts of interest.

Financial Disclosure: The authors declared that this study received no financial support.

Supplementary Table S1, S2: https://d2v96fxpocvxx.cloudfront.net/1dda20c2-8b22-466b-a98e-85ecfb39ceaf/content-images/08fcbe64-8248-4f15-8620-5f1aa9b01ed2.pdf

Supplementary Data S1: https://d2v96fxpocvxx.cloudfront.net/1dda20c2-8b22-466b-a98e-85ecfb39ceaf/content-images/65c87102-2765-44df-9efd-8e8de5049d5c.pdf

References

1. Podder I, Vasudevan B. Dermatological emergencies: what the term encompasses and key features in their diagnosis. In: Verma R, Vasudevan B, editors. Dermatological Emergencies. Boca Raton: CRC Press; 2019. p. 1-12.
2. Chacon AH. Dermatologic emergencies. Cutis. 2015;95:E28-31.
3. Gowda A, Christensen L, Polly S, Barlev D. Necrotizing neutrophilic dermatosis: a diagnostic challenge with a need for multi-disciplinary recognition, a case report. Ann Med Surg (Lond). 2020;57:299-302.
4. Alahmad MS, El-Menyar A, Abdelrahman H, Abdelrahman MA, Aurif F, Shaikh N, et al. Time to diagnose and time to surgery in patients presenting with necrotizing fasciitis: a retrospective analysis. Eur J Trauma Emerg Surg. 2025;51:140.
5. Moola H, Visser WI. Erythroderma in the emergency department: a narrative review. Emerg Care Med. 2026;3:19.
6. Preiksaitis C, Ashenburg N, Bunney G, Chu A, Kabeer R, Riley F, et al. The role of large language models in transforming emergency medicine: scoping review. JMIR Med Inform. 2024;12:e53787.
7. Zhou J, He X, Sun L, Xu J, Chen X, Chu Y, et al. Pre-trained multimodal large language model enhances dermatological diagnosis using SkinGPT-4. Nat Commun. 2024;15:5649.
8. Cirkel L, Lechner F, Henk LA, Krusche M, Hirsch MC, Hertl M, et al. Large language models for dermatological image interpretation: a comparative study. Diagnosis (Berl). 2026;13:75-81.
9. Ju EY, Sinharoy A, Edmonds N, Flowers RH. Diagnostic accuracy of three large language models on clinical images varies by skin tone: findings from the Stanford Diverse Dermatology Images Dataset. Int J Dermatol. 2026;65:1783-5.
10. Cai ZR, Chen ML, Kim J, Novoa RA, Barnes LA, Beam A, et al. Assessment of correctness, content omission, and risk of harm in large language model responses to dermatology continuing medical education questions. J Invest Dermatol. 2024;144:1877-9.
11. Dodiuk-Gad RP, Chung WH, Valeyrie-Allanore L, Shear NH. Stevens–Johnson syndrome and toxic epidermal necrolysis: an update. Am J Clin Dermatol. 2015;16:475-93.
12. Mittal S, Aggarwal Y. Evaluation of large language models in the diagnosis, urgency triage, and initial management of ophthalmic emergencies. Cureus. 2026;18:e101433.
13. Draelos RL, Afreen S, Blasko B, Brazile TL, Chase N, Desai DP, et al. Large language models provide unsafe answers to patient-posed medical questions. NPJ Digit Med. 2026;9:241.
14. Agrawal M, Chen IY, Gulamali F, Joshi S. The evaluation illusion of large language models in medicine. NPJ Digit Med. 2025;8:600.
15. Şen O. Performance of multimodal large language models in misdiagnosed dermatologic cases: a pilot study on diagnostic accuracy and human error replication. Cutan Ocul Toxicol. 2026.
16. Liu X, Duan C, Kim MK, Zhang L, Jee E, Maharjan B, et al. Claude 3 Opus and ChatGPT with GPT-4 in dermoscopic image analysis for melanoma diagnosis: comparative performance analysis. JMIR Med Inform. 2024;12:e59273.
17. Tekchandani N, Mukherjee A, Poonthottam N, Boussios S. Comparative analysis of large language models in dermatological diagnosis: an evaluation of diagnostic accuracy. Cureus. 2025;17:e92089.
18. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc. 2012;19:121-7.
19. Ly DP, Shekelle PG, Song Z. Evidence for anchoring bias during physician decision-making. JAMA Intern Med. 2023;183:818-23.
20. McDuff D, Schaekermann M, Tu T, Palepu A, Wang A, Garrison J, et al. Towards accurate differential diagnosis with large language models. Nature. 2025;642:451-7.
21. Krakowski I, Kim J, Cai ZR, Daneshjou R, Lapins J, Eriksson H, et al. Human–AI interaction in skin cancer diagnosis: a systematic review and meta-analysis. NPJ Digit Med. 2024;7:78.
22. Omiye JA, Gui H, Rezaei SJ, Zou J, Daneshjou R. Large language models in medicine: the potentials and pitfalls: a narrative review. Ann Intern Med. 2024;177:210-20.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.45403
Eurasian J Emerg Med. 2026;25: 467-73

Clinical Characteristics and Factors Associated with Same-encounter Hospital Admission Among Symptomatic Adults Presenting to the Emergency Department After COVID-19

Yılmaz Özgül[1], Muhammed Ediz Sarıhan[1], Talat Kılıç[2]

[1]İnönü University Faculty of Medicine, Department of Emergency Medicine, Malatya, Türkiye
[2]İnönü University Faculty of Medicine, Department of Pulmonology, Malatya, Türkiye

Abstract

Aim: To evaluate adults with previously documented severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection who presented to the emergency department with new or persistent symptoms at least 4 weeks after acute coronavirus disease 2019 (COVID-19).

Materials and Methods: We examined presenting symptoms; demographic and clinical characteristics, laboratory and imaging findings; and factors concurrently associated with all-cause hospital admission during the index post-acute emergency department encounter. Medical records of 760 patients with documented previous SARS-CoV-2 infection who presented to the emergency department between January 2020 and November 2022 were reviewed. Patients with ongoing polymerase chain reaction positivity (n=54), with complaints clearly unrelated to the post-acute presentation (n=74), with incomplete laboratory data (n=84), or aged <18 years (n=48) were excluded. The final cohort comprised 500 patients who presented at least 4 weeks after the documented acute COVID-19 diagnosis.

Results: Of the 500 patients, 275 (55.0%) were female and 225 (45.0%) were male; the median age was 52 years (range, 18-90). Thirty patients had an all-cause hospital admission during the index post-acute emergency department encounter. Admission was more frequent among patients who had required hospital-based care during acute COVID-19 than among those managed as outpatients (19.3% vs. 2.3%; p<0.001). After Holm adjustment, white blood cell count, neutrophil count, blood urea nitrogen, C-reactive protein, procalcitonin (PCT), N-terminal pro-B-type natriuretic peptide, and D-dimer levels were significantly higher among hospitalized patients. In the post hoc exploratory Firth-penalized multivariable model, older age, higher PCT and D-dimer levels, and acute-period hospitalization were associated with same-encounter admission after adjustment.

Conclusion: Older age, higher concurrent PCT and D-dimer levels, and hospitalization during acute COVID-19 were associated with all-cause admission at the index post-acute emergency department encounter. These exploratory same-encounter associations do not establish prediction, causation, or hospitalization attributable to a post-COVID condition, and they require external validation.

Keywords: COVID-19, hospitalization, fatigue, Long COVID, dyspnea

Introduction

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), became a global pandemic with substantial morbidity and mortality (1). Persistent, recurrent, or new symptoms after acute infection remain an important clinical problem and may impair daily functioning (2,3).

The post-acute spectrum is referred to as Long COVID, post-COVID-19 condition, post-acute sequelae of SARS-CoV-2 infection, ongoing symptomatic COVID-19, or post-COVID-19 syndrome. Definitions differ mainly by duration: WHO describes symptoms beginning within 3 months, lasting at least 2 months, and unexplained by another diagnosis (4); CDC describes a chronic condition lasting at least 3 months (2); and NICE uses Long COVID as an umbrella term for symptoms persisting beyond 4 weeks (5).

Corresponding Author: Yılmaz Özgül MD, İnönü University Faculty of Medicine, Department of Emergency, Malatya, Türkiye
E-mail: yilmazozgul259@gmail.com **ORCID ID:** orcid.org/0000-0002-4834-6030

Cite this article as: Özgül Y, Sarıhan ME, Kılıç T. Clinical characteristics and factors associated with Sameencounter Hospital admission among symptomatic adults presenting to the emergency department after COVID-19. Eurasian J Emerg Med. 2026;25: 467-73

Received: 23.06.2026
Accepted: 22.09.2026
Published: 09.10.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Long COVID may follow infection of any severity and fluctuate across organ systems (6). Common complaints include dyspnea, fatigue, and pain (7); proposed mechanisms include immune dysregulation, inflammation, endothelial and microvascular injury, and persistent organ damage (7,8). The diagnosis, treatment, and management of Long COVID are multidisciplinary and symptom-oriented. In addition to routine treatment methods, patient-specific adjunctive therapies are required (8,9).

Long COVID denotes the study's operational threshold of at least 4 weeks after documented infection under the NICE framework (10).

This study examined symptomatic adults who presented to the emergency department at least 4 weeks after documented COVID-19, consistent with the broader NICE framework. We described symptoms, clinical and laboratory findings, and factors that were concurrently associated with all-cause admission during the index post-acute emergency department encounter.

Materials and Methods

This study was approved by the Scientific Research and Publication Ethics Committee of İnönü University, the Republic of Türkiye, under decision number 2022/3020 at the session held on September 20, 2022. The study was a retrospective observational investigation conducted in the emergency department of a tertiary university hospital.

Adults with documented SARS-CoV-2 infection who presented to the emergency department of a tertiary university hospital between January 2020 and November 2022 with new or persistent symptoms at least 4 weeks after diagnosis were eligible. The interval was calculated from diagnosis because recovery dates were inconsistently recorded. Exclusion criteria were ongoing polymerase chain reaction positivity; age <18 years; missing required laboratory measurements, and complaints unrelated to the post-acute presentation (trauma, pregnancy, or forensic cases). Symptom onset and duration were not uniformly recorded; therefore, the cohort was not reclassified using the stricter WHO definition.

Statistical Analysis

SPSS 26.0 was used for descriptive and univariable analyses. Continuous variables were summarized as mean ± standard deviation or median (interquartile range [IQR]), and categorical variables were summarized as n (%). Two groups were compared using independent-samples t-tests or Mann-Whitney U tests; comparisons involving more than two groups used the Kruskal-Wallis test when appropriate. Categorical data were analyzed using the chi-square test, the two-sided Fisher exact test, or the Fisher-Freeman-Halton exact test. Laboratory p values were Holm-adjusted; rank-biserial correlations and 95% confidence intervals (CIs) (5,000 stratified bootstrap resamples) were used to quantify effect size.

Adjusted associations with all-cause ward and/or intensive-care admission during the index post-acute emergency department encounter were examined using Firth-penalized logistic regression. For the revised analysis, a fixed, post hoc, exploratory nine-variable model was defined after a review of the available data and prior analyses, without automated stepwise selection: sex, age, acute-period hospitalization, thoracic involvement, malignancy, white blood cell count (WBC), C-reactive protein (CRP), procalcitonin (PCT), and D-dimer. Female sex and acute outpatient care were the reference categories. Odds ratios (ORs) were reported per 10 years of age, 1 $\times 10^3/\mu L$ WBC, 1 mg/dL CRP, 0.1 ng/mL PCT, and 1 mg/L D-dimer. Profile penalized-likelihood 95% CIs and penalized-likelihood-ratio p-values were reported; variance inflation factors were used to assess collinearity.

Firth estimation was implemented using a custom Python function (Python 3.12; SciPy 1.17) that maximized the Jeffreys-prior-penalized log-likelihood, $\ell F = \ell + 0.5 \log|X'WX|$. Adjusted-score Newton–Raphson updates used step-halving when ℓF decreased. Convergence required a maximum absolute adjusted-score component $<1\times10^{-6}$ and a change in ℓF of (maximum 1,000 iterations). Penalized likelihood-ratio p-values were obtained using constrained refits in which the tested coefficient was fixed at zero. Each coefficient was fixed by profile limits; nuisance parameters were re-optimized with L-BFGS-B; and $2\Delta\ell F = \chi^2_1, 0.95$ was located. Area under the curve (AUC) (scikit-learn 1.8) was estimated using a stratified bootstrap 95% CI (5,000 resamples); calibration was assessed using intercept, slope, and Brier score. The optimism-adjusted AUC was estimated using 1,000 bootstrap resamples (seed 20260908), refitting the fixed model in each sample. CRP values <0.33 mg/dL were set to 0.33 mg/dL for analysis. PCT was standardized to ng/mL; eight misplaced decimal separators were corrected after review of the source data.

All tests were two-sided, with $p<0.05$. Complete laboratory and model data were available for all 500 included patients; 84 screened patients lacked at least one required measurement. It was not possible to determine whether missingness was related to severity or disposition because outcomes were incomplete for excluded records. Patient selection is shown in Figure 1.

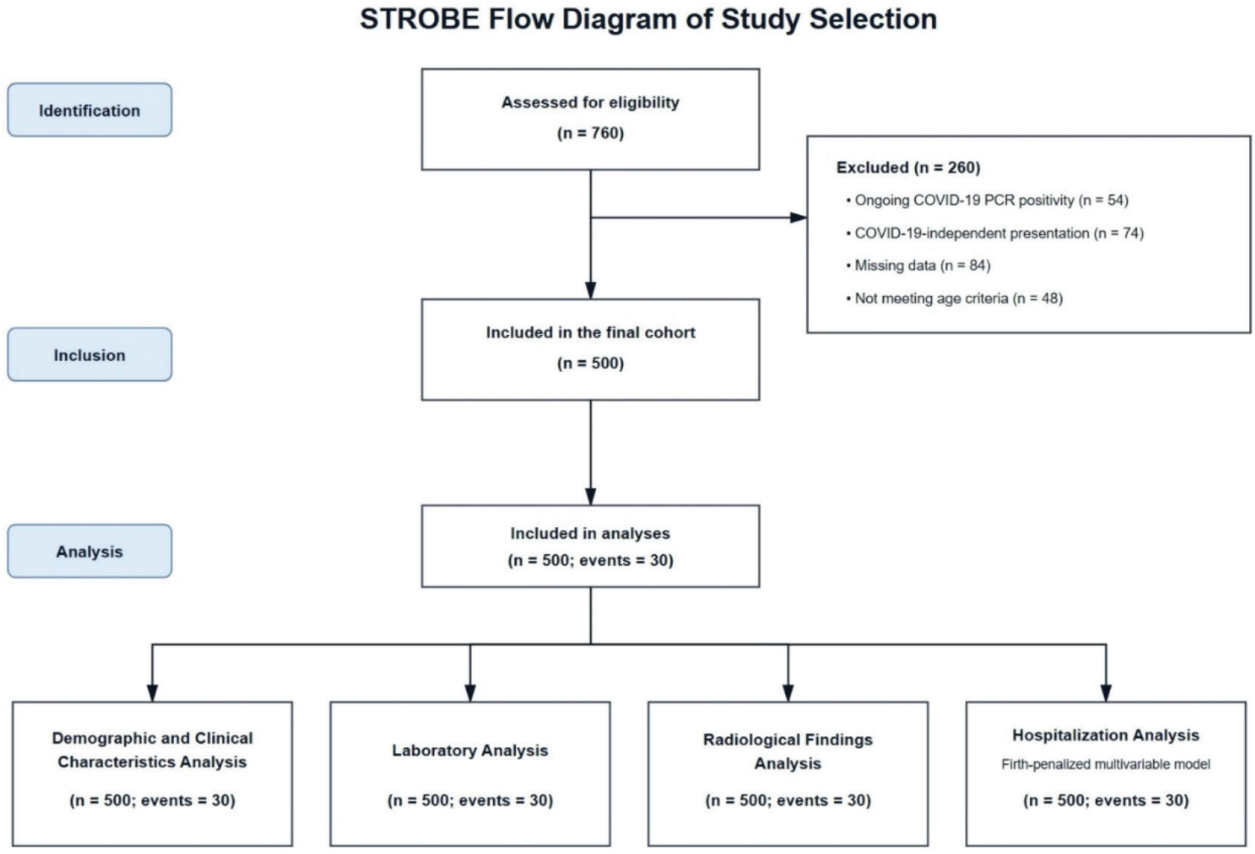


Figure 1. STROBE flow diagram of patient selection and study analysis
COVID-19: Coronavirus disease 2019, PCR: Polymerase chain reaction

Results

Among 500 patients, 275 (55.0%) were women and 225 (45.0%) were men; the median age was 52 years (range, 18-90). The median time from acute diagnosis to presentation was 54 days (IQR, 40-69; range, 30-139). Of these, 445 (89.0%) presented at 4-12 weeks and 55 (11.0%) thereafter. Acute care was outpatient for 391 patients (78.2%), ward-based for 76 (15.2%), and intensive care for 33 (6.6%). At least one comorbidity was present in 71.6% of patients, and thoracic involvement was present in 377 (75.4%) (Table 1).

Common symptoms were dyspnea (73.8%), weakness or fatigue (68.0%), exertional dyspnea (53.4%), loss of appetite (47.0%), cough (33.4%), chest pain (32.2%), myalgia (23.8%), sputum (20.4%), and brain fog (8.0%). Symptoms were not mutually exclusive.

Thirty patients were admitted at the index encounter: 19 to wards and 11 to intensive care. Recorded diagnoses that were not mutually exclusive included pulmonary embolism and pneumonia (n=11 each), residual fibrotic changes (n=6), heart failure and gastrointestinal bleeding (n=2 each), and myocardial infarction, total atelectasis, elevated liver enzymes,

Table 1. Descriptive statistics of patients' sociodemographic characteristics, comorbidities, and acute COVID-19 care

	Number (n)	Percent (%)
Sex		
Female	275	55.0
Male	225	45.0
Acute COVID-19 care		
Outpatient	391	78.2
Inpatient ward	76	15.2
Intensive care unit	33	6.6
Thoracic involvement		
Present	377	75.4
Absent	123	24.6
Comorbidities		
CAD	95	19.0
HT	152	30.4
DM	68	13.6
Malignancy	34	6.8
Asthma/COPD	78	15.6

CAD: Coronary artery disease, HT: Hypertension, DM: Diabetes mellitus, COPD: Chronic obstructive pulmonary disease, COVID-19: Coronavirus disease 2019

adrenal insufficiency, alveolar hemorrhage, and diabetes-related decompensation (n=1 each). Diagnoses and routine-care admission decisions were not independently adjudicated.

Acute and post-acute care levels were associated (Fisher-Freeman-Halton p<0.001). Specifically, 5/11 post-acute intensive-care admissions required acute intensive care, and 12/19 post-acute ward admissions received acute ward care (Table 2).

After Holm correction, admitted patients had higher WBC (9.87 vs 7.60 x10^3/uL; adjusted p=0.012), neutrophils (5.97 vs 4.38 x10^3/uL; p=0.019), blood urea nitrogen (18.45 vs 14.49 mg/dL; p=0.009), CRP (2.96 vs <0.33 mg/dL), PCT (0.099 vs 0.034 ng/mL), N-terminal pro-B-type natriuretic peptide (260.5 vs 52 pg/mL), and D-dimer (1.05 vs 0.33 mg/L FEU; the latter four p<0.001) (Table 3). Glucose and alanine aminotransferase were no longer significant after correction; other laboratory measures did not differ.

The fixed post hoc nine-variable Firth model included all 500 patients and 30 events (Table 4). Admission was associated with age (adjusted OR per 10 years=1.45; 95% profile penalized-

Table 2. Comparison of acute COVID-19 care and disposition at the index post-acute emergency department encounter

		Index post-acute ED disposition				
Acute COVID-19 care		**Outpatient**	**Ward**	**ICU**	**Total**	**P**
	Outpatient	382 (97.7%)	5 (1.3%)	4 (1.0%)	391	
	Ward	62 (81.6%)	12 (15.8%)	2 (2.6%)	76	**<0.001**
	ICU	26 (78.8%)	2 (6.1%)	5 (15.2%)	33	
Total		470 (94.0%)	19 (3.8%)	11 (2.2%)	500	

Data are presented as n (row percentage). The p value was calculated using the Fisher-Freeman-Halton exact test. Percentages reported in the text for index admissions to post-acute wards and to intensive care were calculated using the corresponding disposition category as the denominator. ICU: Intensive care unit, ED: Emergency department, COVID-19: Coronavirus disease 2019

Table 3. Comparison of laboratory parameters by disposition at the index post-acute emergency department encounter

Parameter	Reference range	Outpatient, median (IQR)	Hospitalized, median (IQR)	Effect size, r_rb. (95% CI)	p	Holm-adjusted p
WBC (×10^3/μL)	4.0-10.0	7.60 (6.20-9.36)	9.87 (7.32-11.81)	0.353 (0.132-0.556)	0.001	**0.012**
Hemoglobin (g/dL)	Male: 13.5-17.5; female: 12.0-16.0	13.9 (12.4-15.2)	13.2 (10.7-14.9)	-0.176 (-0.408-0.065)	0.107	0.639
Hematocrit (%)	Male: 40-52; female: 36-47	42.7 (38.9-46.2)	40.5 (34.0-45.9)	-0.172 (-0.409-0.075)	0.114	0.639
Platelet (×10^3/μL)	150-400	266 (222-310)	292 (223-351)	0.144 (-0.103-0.391)	0.187	0.639
Lymphocyte (×10^3/μL)	1.0-4.0	2.23 (1.73-2.79)	2.08 (1.37-2.67)	-0.157 (-0.381-0.073)	0.149	0.639
Neutrophil (×10^3/μL)	1.5-7.5	4.38 (3.34-5.88)	5.97 (3.73-8.37)	0.335 (0.104-0.537)	0.002	**0.019**
Glucose (mg/dL)	70-99 (fasting)	99 (90-115)	111 (98-142)	0.246 (0.039–0.447)	0.024	0.165
AST (U/L)	10-40	21.0 (17.0-25.0)	24.5 (17.0-33.0)	0.159 (-0.096-0.412)	0.144	0.639
ALT (U/L)	7-56	19.0 (14.0-28.0)	25.5 (18.8-37.8)	0.262 (0.050-0.464)	0.016	0.128
BUN (mg/dL)	7-20	14.49 (11.21-18.22)	18.45 (14.14-25.58)	0.366 (0.165–0.549)	<0.001	**0.009**
Creatinine (mg/dL)	Male: 0.74-1.35; female: 0.59-1.04	0.81 (0.70-1.00)	0.95 (0.72-1.12)	0.166 (-0.085-0.407)	0.123	0.639
CRP (mg/dL)	0-0.33	<0.33 (<0.33-0.69)	2.96 (1.39-5.69)	0.644 (0.463-0.796)	<0.001	**<0.001**
PCT (ng/mL)	<0.05	0.034 (0.023-0.061)	0.099 (0.049-0.307)	0.557 (0.356–0.730)	<0.001	**<0.001**
NT-proBNP (pg/mL)	<125 (<75 y); <450 (≥75 y)	52 (<20-173)	260.5 (137.0-1072.5)	0.499 (0.302-0.681)	<0.001	**<0.001**
D-dimer (mg/L FEU)	≤0.50	0.33 (0.21-0.54)	1.05 (0.48-7.04)	0.591 (0.409-0.755)	<0.001	**<0.001**

Continuous variables are presented as median (interquartile range) and compared using two-sided Mann–Whitney U tests. All parameters were available for 470 outpatients and 30 hospitalized patients (total n=500). Effect sizes are unadjusted rank-biserial correlations (r_rb.); positive values indicate higher measurements among hospitalized patients. The 95% CIs were estimated using 5,000 stratified bootstrap resamples. Holm-adjusted p-values account for multiple testing across laboratory parameters, with statistical significance defined as adjusted p<0.05. C-reactive protein values reported below the analytical limit (<0.33 mg/dL) were summarized using the reported-limit notation and were assigned a value of 0.33 mg/dL for statistical analysis. Hospitalization includes ward and intensive care admissions during the index encounter.
ALT: Alanine aminotransferase, AST: Aspartate aminotransferase, BUN: Blood urea nitrogen, CI: Confidence interval, CRP: C-reactive protein, FEU: Fibrinogen-equivalent units, IQR: Interquartile range, NT-proBNP: N-terminal pro-B-type natriuretic peptide, PCT: Procalcitonin, WBC: White blood cell count

likelihood CI, 1.07-2.01; p=0.017), PCT (per 0.1 ng/mL=1.12; 1.02-1.24; p=0.013), D-dimer (per 1 mg/L=1.26; 1.10-1.49; p<0.001), and acute-period hospitalization (OR=3.45; 1.31-9.39; p=0.012). Concurrent biomarkers should not be interpreted as prospective predictors.

Sex, thoracic involvement, malignancy, and CRP were not significant after adjustment. WBC was borderline (adjusted OR per 1 $\times 10^3/\mu L$=1.117; 95% profile penalized-likelihood CI= 0.9999-1.2386; p=0.0502). Apparent AUC was 0.914 (bootstrap 95% CI, 0.870–0.951); optimism-adjusted AUC was 0.885. Calibration intercept, slope, and Brier score were −0.098, 1.078, and 0.039, respectively; VIFs ranged from 1.06 to 1.42.

Unadjusted analyses associated admission with male sex (OR=3.04; 95% CI, 1.36-6.78; p=0.007) and the presence of malignancy (OR=3.04; 1.08-8.53; p=0.044), but not with coronary artery disease, asthma/chronic obstructive pulmonary disease, diabetes, or hypertension (Table 5).

After excluding isolated myocardial infarction and gastrointestinal bleeding, 498 patients and 28 events remained. Estimates

Table 4. Post hoc exploratory firth-penalized multivariable analysis of all-cause admission at the index post-acute emergency department encounter

Variable	β	Adjusted OR	95% CI	p
Age, per 10-year increase	0.37	1.45	1.07-2.01	**0.017**
Thoracic involvement	−0.05	0.96	0.31-3.35	0.939
Malignancy	0.39	1.48	0.36-4.95	0.558
WBC, per 1 $\times 10^3/\mu L$ increase	0.110	1.117	0.9999-1.2386	0.0502
CRP, per 1 mg/dL increase	0.04	1.04	0.94-1.18	0.414
PCT, per 0.1 ng/mL increase	0.11	1.12	1.02-1.24	**0.013**
D-dimer, per 1 mg/L increase	0.23	1.26	1.10-1.49	**<0.001**
Acute-period hospitalization (yes vs outpatient)	1.24	3.45	1.31-9.39	**0.012**
Male sex (vs female)	0.73	2.08	0.83-5.57	0.117

OR: Odds ratio, CI: Confidence interval, WBC: White blood cell count, CRP: C-reactive protein, PCT: Procalcitonin. Hospitalization was defined as all-cause admission to a ward and/or an intensive care unit during the index post-acute emergency department encounter. The fixed post hoc exploratory model included nine variables, 30 events, and all 500 patients. Female sex and outpatient acute care were the reference categories. ORs are reported per 10-year increase in age, per 1 × $10^3/\mu L$ increase in WBC, per 1 mg/dL increase in CRP, per 0.1 ng/mL increase in PCT, and per 1 mg/L increase in D-dimer. Profile penalized-likelihood 95% CIs and penalized likelihood-ratio p values were obtained from constrained refits, and therefore used the same inferential framework. The apparent AUC was 0.914 (bootstrap 95% CI, 0.870-0.951); the optimism-adjusted AUC after 1,000 bootstrap resamples was 0.885. The calibration intercept was −0.098, the slope was 1.078, and the Brier score was 0.039. Statistical significance was defined as p<0.05

Table 5. Unadjusted association of sex and comorbidities with all-cause admission at the index post-acute emergency department encounter

Variable	Status	Outpatient (n)	Hospitalized (n)	Total (n)	Crude OR (95% CI)	p
Sex	Female	266	9	275	Reference	**0.007**
	Male	204	21	225	**3.04 (1.36-6.78)**	
CAD	Absent	385	20	405	Reference	0.052
	Present	85	10	95	2.26 (1.02-5.01)	
Asthma/COPD	Absent	396	26	422	Reference	1.000
	Present	74	4	78	0.82 (0.28-2.41)	
Malignancy	Absent	441	25	466	Reference	**0.044**
	Present	29	5	34	**3.04 (1.08-8.53)**	
DM	Absent	406	26	432	Reference	1.000
	Present	64	4	68	0.98 (0.33-2.89)	
HT	Absent	330	18	348	Reference	0.305
	Present	140	12	152	1.57 (0.74-3.34)	

Data are presented as numbers. Crude ORs and Wald-type 95% CIs indicate unadjusted odds of hospitalization; the OR for sex compares male patients with female patients. Female sex and the absence of the relevant comorbidity were the reference categories. P values were calculated using two-sided Fisher's exact tests. CAD: Coronary artery disease, COPD: Chronic obstructive pulmonary disease, DM: Diabetes mellitus, HT: Hypertension. Crude odds ratios and Wald-type 95% confidence intervals are reported as measures of association. Because of sparse cell counts, two-sided Fisher's exact tests were used to calculate p values. Consequently, confidence intervals and p-values may not correspond exactly because they were obtained using different inferential methods

were similar: adjusted ORs were 1.38 per 10 years (95% profile CI, 1.0001-1.95; p=0.0499), 1.11 per 0.1 ng/mL PCT (1.02-1.23; p=0.015), 1.27 per 1 mg/L D-dimer (1.10-1.50; p<0.001), and 3.85 for acute hospitalization (1.41-11.00; p=0.009). The apparent AUC was 0.918 (bootstrap 95% CI, 0.873-0.955).

Discussion

Long COVID is thought to be associated with female sex, advanced age, obesity, diabetes, cardiovascular diseases, and other comorbidities. In our study population, however, male sex and the presence of malignancy were associated with hospitalization only in unadjusted analyses and were not significant after adjustment, highlighting the need for cautious interpretation (11,12).

Literature reviews include studies showing that hospitalization during the acute phase is associated with hospitalization during the long COVID period (13,14). Consistent with reports of a greater post-acute burden following severe acute illness, acute-phase hospitalization remained associated with hospitalization during the initial post-acute emergency department encounter (adjusted OR=3.45). This may reflect persistent frailty or healthcare needs; however, the observational design and the all-cause nature of the outcome preclude attributing the findings to causal relationships or specific post-COVID conditions.

Inflammation and endothelial dysfunction are factors potentially contributing to post-acute thromboembolic and multisystem symptoms; higher levels of inflammatory, cardiac stress, and coagulation markers have been observed in patients hospitalized following post-acute presentations (15,16). In the exploratory model, advanced age, concurrent PCT and D-dimer levels, and acute-phase hospitalization showed a significant association with hospitalization during the prolonged period, while WBC count was of borderline significance. Laboratory parameters may reflect acute conditions or influence clinical decisions. Therefore, these biomarkers should be interpreted not as predictors of future deterioration, diagnostic biomarkers, or causal determinants of Long COVID severity, but rather as indicators associated concurrently with hospitalization.

Study Limitations

The retrospective, single-center design relied on routine records and included only symptomatic emergency patients, limiting generalizability and permitting selection, referral, measurement, and classification bias. Symptom onset, duration, and recovery dates were inconsistently documented; therefore, the cohort follows a broad NICE-based threshold. Excluding 84 records with incomplete laboratory data may have favored patients who were more extensively tested. Vaccination, variant, smoking, body mass index, acute treatments and respiratory support, length of stay, readmission, mortality, and socioeconomic status were unavailable or incomplete, thereby leaving residual confounding.

Only 30 events supported nine model variables. Firth penalization and bootstrap validation reduce but do not eliminate overfitting. Some CIs were wide, and external validation was unavailable. The model was specified post hoc during revision after review of the available data and prior analyses, although no automated stepwise selection was used. Refitting the fixed, selected model in each bootstrap sample estimates optimism conditional on that model, but does not fully account for optimism introduced by analyst-guided model specification. The all-cause outcome was based on routine clinician decisions without standardized criteria or independent adjudication of diagnoses. Alternative acute diagnoses and uncertainty regarding causation remain. Reports of pulmonary fibrosis reflected residual or preliminary imaging interpretations rather than standardized longitudinal diagnoses. Findings are exploratory, not prognostic.

Conclusion

Among symptomatic adults presenting at least 4 weeks after COVID-19, older age, concurrent PCT and D-dimer levels, and hospitalization during acute infection were associated after adjustment with all-cause admission at the index post-acute emergency department encounter. These exploratory same-encounter associations are neither causal nor specific to a post-COVID condition. Prospective multicenter studies with standardized definitions, adjudicated diagnoses, and longitudinal outcomes are needed.

Ethics

Ethics Committee Approval: This study was approved by the Scientific Research and Publication Ethics Committee of İnönü University, the Republic of Türkiye, under decision number 2022/3020 at the session held on September 20, 2022.

Informed Consent: The study was a retrospective observational investigation conducted in the emergency department of a tertiary university hospital.

Footnotes

Authorship Contributions

Surgical and Medical Practices: Y.Ö., Concept: M.E.S., T.K., Design: M.E.S., Data Collection or Processing: Y.Ö., Analysis or Interpretation: M.E.S., T.K., Literature Search: Y.Ö., Writing: Y.Ö.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. World Health Organization. WHO COVID-19 Dashboard [Internet]. Geneva: World Health Organization Available from: https://data.who.int/dashboards/covid19/summary
2. Centers for Disease Control and Prevention. Long COVID: Basics [Internet]. Atlanta (GA): Centers for Disease Control and Prevention. Available from: https://www.cdc.gov/long-covid/about/index.html
3. National Institute for Health and Care Excellence. COVID-19 rapid guideline: managing the long-term effects of COVID-19. NICE guideline NG188 [Internet]. London: NICE; 2020. Available from: https://www.nice.org.uk/guidance/ng188
4. World Health Organization. Post COVID-19 condition (Long COVID) [Internet]. Geneva: World Health Organization. Available from: https://www.who.int/news-room/fact-sheets/detail/post-covid-19-condition-(long-covid)
5. Venkatesan P. NICE guideline on long COVID. Lancet Respir Med. 2021;9:129.
6. Donald J, Bilasy SE, Yang C, El-Shamy A. Exploring the complexities of Long COVID. Viruses. 2024;16:1060.
7. Taribagil P, Creer D, Tahir H. 'Long COVID' syndrome. BMJ Case Rep. 2021;14:e241485.
8. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. Nat Rev Microbiol. 2023;21:133-46.
9. Cheng AL, Herman E, Abramoff B, Anderson JR, Azola A, Baratta JM, et al. Multidisciplinary collaborative guidance on the assessment and treatment of patients with Long COVID: A compendium statement. PM R. 2025;17:684-708.
10. Shah W, Hillman T, Playford ED, Hishmeh L. Managing the long term effects of covid-19: summary of NICE, SIGN, and RCGP rapid guideline. BMJ. 2021;372:n136.
11. Ashmawy R, Hammouda EA, El-Maradny YA, Aboelsaad I, Hussein M, Uversky VN, et al. Interplay between comorbidities and long COVID: challenges and multidisciplinary approaches. Biomolecules. 2024;14:835.
12. Huang C, Huang L, Wang Y, Li X, Ren L, Gu X, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. Lancet. 2021;397:220-32.
13. Nalbandian A, Sehgal K, Gupta A, Madhavan MV, McGroder C, Stevens JS, et al. Post-acute COVID-19 syndrome. Nat Med. 2021;27:601-15.
14. Moreno-Pérez O, Merino E, Leon-Ramirez JM, Andres M, Ramos JM, Arenas-Jiménez J, et al. Post-acute COVID-19 syndrome. Incidence and risk factors: a Mediterranean cohort study. J Infect. 2021;82:378-83.
15. Mandal S, Barnett J, Brill SE, Brown JS, Denneny EK, Hare SS, et al. Long-COVID: a cross-sectional study of persisting symptoms, biomarker and imaging abnormalities following hospitalisation for COVID-19. Thorax. 2021;76:396-8.
16. Tsilingiris D, Vallianou NG, Karampela I, Christodoulatos GS, Papavasileiou G, Petropoulou D, et al. Laboratory findings and biomarkers in long COVID: what do we know so far? Insights into epidemiology, pathogenesis, therapeutic perspectives and challenges. Int J Mol Sci. 2023;24:10458.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.98958
Eurasian J Emerg Med. 2026;25: 474-80

Comparison of Clinical, Laboratory, and Inflammatory Parameters in Pediatric Upper and Lower Gastrointestinal Bleeding

Süleyman Nogay[1], Nuriye Aslı Melekoğlu[2]

[1]Malatya Turgut Özal University Faculty of Medicine, Department of Emergency Medicine, Malatya, Türkiye
[2]Malatya Turgut Özal University Faculty of Medicine, Department of Pediatrics, Malatya, Türkiye

Abstract

Aim: Pediatric gastrointestinal (GI) bleeding is an important emergency ranging from mild, self-limiting conditions to severe blood loss. Although laboratory parameters and inflammatory indices have been investigated for predicting clinical course, pediatric data remain limited.

Materials and Methods: This retrospective study included 29 pediatric patients hospitalized for GI bleeding. Patients were classified as having upper or lower GI bleeding. Hematological and biochemical parameters, neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), systemic immune-inflammation index (SII) and systemic inflammation response index (SIRI), and urea/creatinine (Cr) ratio were evaluated. Groups were compared, and associations with hospital stay were assessed by Spearman correlation analysis.

Results: Upper GI bleeding was identified in 20 (69%) patients, and lower GI bleeding was identified in 9 (31%). Hemoglobin and hematocrit were lower, and aspartate aminotransferase was higher, in the upper GI group, but these differences were not statistically significant ($p=0.052$, $p=0.089$, and $p=0.054$, respectively). NLR, PLR, SII, and SIRI did not differ between groups. Hospital stay was negatively correlated with Cr, albumin, calcium, hemoglobin, and hematocrit, and positively correlated with white blood cell, neutrophil, and SIRI. No significant correlations were found with NLR, PLR, or SII.

Conclusion: Routine laboratory parameters and inflammatory indices alone appear insufficient to distinguish upper from lower GI bleeding. However, the association of SIRI with hospital stay represents a preliminary finding, suggesting its potential complementary role in evaluating the clinical course. Larger multicenter prospective studies are needed to validate this finding.

Keywords: Gastrointestinal hemorrhage, child, inflammation, biomarkers, length of stay

Introduction

Gastrointestinal (GI) bleeding is an important pediatric emergency, ranging from mild, self-limiting bleeding to severe conditions causing hemodynamic instability and hemorrhagic shock (1,2). Although approximately 80% of pediatric GI bleeding resolves spontaneously, early identification of patients at risk of severe bleeding is essential to reduce morbidity and complications (1). Assessment of hemodynamic status and bleeding severity, identification of the potential bleeding source, and early recognition of patients requiring advanced diagnostic or therapeutic interventions are fundamental to emergency department management (2,3).

GI bleeding is classified as upper or lower according to anatomical location. Upper GI bleeding commonly presents with hematemesis or melena, whereas hematochezia is more frequent in lower GI bleeding (1,2). Etiology varies considerably with age, comorbidities, and bleeding location, complicating diagnosis in children (1,4). Particularly in upper GI bleeding, assessing severity and determining the need and appropriate timing for endoscopy are essential for clinical management (5,6). Endoscopy also enables therapeutic intervention when required (7,8).

Routine laboratory markers, including hemoglobin (HGB), hematocrit (HCT), platelet (PLT), urea, creatinine (Cr), albumin (ALB), C-reactive protein (CRP), and coagulation parameters, are used in the initial evaluation of GI bleeding (2,3,6). Hematological and composite inflammatory indices derived from routine blood counts have also been investigated in acute clinical conditions (9,10). Neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), systemic immune-inflammation index (SII) and systemic inflammation response index (SIRI) are inexpensive and

Corresponding Author: Asst. Prof. Süleyman Nogay, Malatya Turgut Özal University Faculty of Medicine, Department of Emergency Medicine, Malatya, Türkiye
E-mail: drsnogay@gmail.com **ORCID ID:** orcid.org/0000-0003-1184-9467

Cite this article as: Nogay S, Melekoğlu NA. Comparison of clinical, laboratory, and inflammatory parameters in pediatric upper and lower gastrointestinal bleeding. Eurasian J Emerg Med. 2026;25: 474-80.

Received: 13.09.2026
Accepted: 07.10.2026
Published: 20.10.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

accessible markers reflecting systemic inflammatory and immune responses. In acute upper GI bleeding, inflammatory indices have been associated with in-hospital mortality (10). However, age-specific reference ranges for SII, NLR, and PLR in healthy children emphasize the importance of age in interpreting these markers (11). Evidence regarding their clinical significance and potential utility in pediatric GI bleeding nevertheless remains limited.

The aim of this study was to compare demographic characteristics, routine laboratory parameters, and inflammatory indices, including NLR, PLR, SII, and SIRI, between pediatric patients with upper GI bleeding and those with lower GI bleeding, and to assess their associations with length of hospital stay.

Materials and Methods

Study Design and Setting

Following ethics approval by the Malatya Turgut Özal University Health Sciences Scientific Research Ethics Committee (approval date: 11.09.2026; approval no: E-30785963-020-421429), the data from pediatric patients who were admitted to the Malatya Training and Research Hospital emergency department between January and December 2025 and hospitalized with suspected GI bleeding were retrospectively reviewed.

Laboratory parameters potentially associated with clinical course and hospital stay were evaluated: complete blood count parameters [white blood cell (WBC), HGB, HCT, PLT, lymphocyte (LYM), monocyte (MONO), and neutrophil (NEU)]; biochemical parameters [CRP, urea, Cr, aspartate aminotransferase (AST), alanine aminotransferase (ALT), ALB, sodium (Na), potassium (K), and calcium (Ca)]; international normalized ratio (INR); derived ratios [NLR, PLR, and urea-to-creatinine ratio (urea/Cr)]; and inflammatory indices (SII and SIRI).

Composite hematological and inflammatory indices were calculated from admission laboratory values. NLR was calculated as NEU/LYM, and PLR was calculated as PLT/LYM. SII was calculated as PLT × NEU/LYM and SIRI was calculated as NEU × MONO/LYM. The urea/Cr ratio was calculated by dividing urea by creatinine.

GI bleeding was anatomically classified as upper or lower. Bleeding proximal to the ligament of Treitz was considered upper GI bleeding, and distal bleeding as lower GI bleeding (1).

Selection of Participants

Using the Malatya Training and Research Hospital Information Management System, 38 patients who were recorded between January and December 2025 with ICD-10 codes K92.0 (hematemesis), K92.1 (melena), K92.2 (unspecified GI bleeding), or K62.5 (hemorrhage of anus and rectum) were identified. Electronic medical records were reviewed for demographic information, admission laboratory results, endoscopic and colonoscopic findings, clinical diagnoses, and hospital stay. Nine patients who did not meet the inclusion criteria were excluded, and data from 29 eligible pediatric patients were retrospectively analyzed.

Inclusion Criteria

- Pediatric patients aged <18 years,
- Patients hospitalized from the emergency department with suspected GI bleeding,
- Patients with complete data available in the hospital information management system,
- Patients admitted between January and December 2025.

Exclusion Criteria

- Patients aged ≥18 years,
- Patients with incomplete data in the Hospital information management system,
- Patients transferred from the emergency department to other hospitals or who refused hospitalization and left the hospital.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, version 27.0 (IBM Corp., Armonk, NY, USA). Normality of continuous variables was assessed using the Shapiro-Wilk test, visual inspection of histograms and Q-Q plots, and evaluation of skewness and kurtosis. Normally distributed variables were expressed as mean ± standard deviation (SD), non-normally distributed variables as median and interquartile range [25^{th}-75^{th} percentile (IQR)], and categorical variables as number (n) and percentage (%).

Upper and lower GI bleeding groups were compared using the independent-samples t-test for normally distributed variables and the Mann-Whitney U test for non-normally distributed variables. Associations between laboratory parameters, inflammatory indices, and hospital stay were assessed using Spearman rank correlation analysis. Correlation coefficients were reported as Spearman's rho (ρ). All statistical tests were two-tailed, and $p<0.05$ was considered statistically significant.

Results

A total of 29 pediatric patients were included. The median age was 14 years (IQR: 8-15; range: 1-17 years); 14 (48.3%) were female and 15 (51.7%) were male. Upper GI bleeding (esophagitis, gastritis,

gastric ulcer, bulbitis) occurred in 20 patients (69.0%), and lower GI bleeding (ileitis, colitis, hemorrhoids) occurred in 9 patients (31.0%). The median hospital stay was 1 day (IQR: 1-3; range: 1-8 days). Endoscopy alone was performed in 20 (69.0%) patients, while 9 (31.0%) underwent both endoscopy and colonoscopy. Admission laboratory parameters and inflammatory indices were evaluated.

Demographic, laboratory, and inflammatory parameters are presented in Table 1 as mean ± SD for normally distributed variables and median (25^{th}-75^{th} percentile) for non-normally distributed variables.

Parametric laboratory data from patients with upper GI bleeding and from patients with lower GI bleeding were compared using the independent-samples t-test. HGB and HCT levels were lower in the upper GI bleeding group, but the differences were not statistically significant (p=0.052 and p=0.089, respectively). AST levels were higher in the upper GI bleeding group, but the difference was not statistically significant (p=0.054). The urea/Cr ratio did not differ significantly between the upper and lower GI bleeding groups (p=0.308). No significant differences were found in other parametric laboratory variables (all p>0.05) (Table 2).

Table 1. Descriptive statistics of study variables presented as mean ± SD or median (Q1-Q3)

Parameters	Mean ± SD	Median (Q1-Q3)
Cr (mg/dL)	0.59±0.12	—
Urea (mg/dL)	23.82±7.26	—
AST (U/L)	25.93±11.49	—
ALB (g/dL)	4.45±0.46	—
K (mmol/L)	4.32±0.43	—
Ca (mg/dL)	9.58±0.54	—
WBC ($\times10^9$/L)	8.98±2.41	—
PLT ($\times10^9$/L)	308.83±77.11	—
HGB (g/dL)	12.71±2.22	—
HCT (%)	38.93±6.07	—
INR	1.06±0.08	—
Urea/Cr	41.96±14.24	—
Age (years)	—	14 (8-15)
ALT (U/L)	—	15 (11-20)
Na (mmol/L)	—	138 (136-140)
CRP (mg/L)	—	1.00 (0.60-2.60)
NEU ($\times10^9$/L)	—	4.90 (3.82-6.46)
LYM ($\times10^9$/L)	—	2.55 (2.03-2.96)
MONO ($\times10^9$/L)	—	0.58 (0.51-0.70)
NLR	—	1.95 (1.22-2.98)
PLR	—	125.30 (85.05-160.20)
SII	—	559.20 (343.60-1071.45)
SIRI	—	1.22 (0.73-1.99)
Hospital stay (days)	—	1 (1-3)

Q1 and Q3 indicate the 25^{th} and 75^{th} percentiles, respectively
SD: Standard deviation, Cr: Creatinine, AST: Aspartate aminotransferase, ALB: Albumin, K: Potassium, Ca: Calcium, WBC: White blood cell, PLT: Platelet, HGB: Hemoglobin, HCT: Hematocrit, INR: International normalized ratio, ALT: Alanine aminotransferase, Na: Sodium, CRP: C-reactive protein, NEU: Neutrophil, LYM: Lymphocyte, MONO: Monocyte, NLR: Neutrophil-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, SII: Systemic immune-inflammation index, SIRI: Systemic inflammation response index

The Mann-Whitney U test was used to compare non-normally distributed variables between upper and lower GI bleeding groups. No statistically significant differences were observed for age, ALT, Na, CRP, NEU, LYM, MONO, NLR, PLR, SII, or SIRI (all p>0.05) (Table 3).

Associations between hospital stay, laboratory parameters, and inflammatory indices were assessed using Spearman correlation analysis. Hospital stay showed significant, moderate negative correlations with Cr (ρ=-0.531; p=0.003), ALB (ρ=-0.616; p<0.001), Ca (ρ=-0.516; p=0.004), HGB (ρ=-0.449; p=0.015), and HCT (ρ=-0.562; p=0.002). Significant moderate positive correlations were found with WBC (ρ=0.408; p=0.028), NEU (ρ=0.511; p=0.005), and SIRI (ρ=0.450; p=0.014). No significant correlations were observed with urea, AST, ALT, Na, K, CRP, PLT, LYM, MONO, INR, NLR, PLR, SII, or urea/Cr ratio (all p>0.05) (Table 4).

Discussion

Pediatric GI bleeding has a variable clinical course, making early risk stratification important. Evidence regarding readily available laboratory markers for predicting severity and clinical outcomes remains limited, particularly in children (3,12). We evaluated routine laboratory parameters and inflammatory indices, including NLR, PLR, SII, and SIRI, in pediatric patients with upper and lower GI bleeding. Although inflammatory indices did not differ significantly between groups, several parameters were associated with hospital stay. SIRI correlated positively with hospital stay, whereas HGB, HCT, ALB, Ca, and Cr correlated negatively. These findings suggest that laboratory parameters and inflammatory indices may provide complementary information regarding the clinical course.

HGB and HCT levels were lower in upper than in lower GI bleeding, although the differences were not statistically significant (p=0.052 and p=0.089, respectively). This may reflect greater blood loss in some upper GI bleeding cases. However, normal HGB values early after acute bleeding do not exclude significant blood loss, emphasizing serial measurements alongside clinical assessment (12). In 55 children with melena, HGB ≥3 g/dL below the age-adjusted lower limit was associated

Table 2. Comparison of parametric laboratory parameters between upper and lower gastrointestinal bleeding groups

Parameters	Upper GIB (n=20) Mean ± SD	Lower GIB (n=9) Mean ± SD	p-value*	Effect size (Hedges' g)
Cr (mg/dL)	0.58±0.12	0.60±0.09	0.687	0.159
Urea (mg/dL)	24.55±8.02	22.21±5.19	0.432	0.311
AST (U/L)	28.35±12.12	20.55±8.11	0.054	0.683
ALB (g/dL)	4.36±0.43	4.64±0.47	0.131	0.607
K (mmol/L)	4.35±0.39	4.23±0.51	0.528	0.249
Ca (mg/dL)	9.57±0.55	9.60±0.52	0.810	0.044
WBC ($\times10^9$/L)	8.65±2.56	9.71±1.94	0.280	0.430
PLT ($\times10^9$/L)	305.10±86.90	317.11±52.56	0.705	0.149
HGB (g/dL)	12.17±2.20	13.90±1.86	0.052	0.795
HCT (%)	37.64±6.18	41.79±4.98	0.089	0.689
INR	1.06±0.08	1.05±0.08	0.700	0.044
Urea/Cr	43.79±15.34	37.86±11.09	0.308	0.405

*: Independent samples t-test, Cr: Creatinine, GIB: Gastrointestinal bleeding, SD: Standard deviation, AST: Aspartate aminotransferase, ALB: Albumin, K: Potassium, Ca: Calcium, WBC: White blood cell, PLT: Platelet, HGB: Hemoglobin, HCT: Hematocrit, INR: International normalized ratio, Urea/Cr: Urea-to-creatinine ratio

Table 3. Comparison of non-parametric laboratory parameters and inflammatory indices between upper and lower gastrointestinal bleeding groups

Parameters	GIB	n	Median	IQR (25^{th}-75^{th} percentile)	p-value*	Effect size (r)**
Age (years)	Upper	20	15	7.75-15	0.519	0.120
	Lower	9	13	9-14		
ALT (U/L)	Upper	20	13	10.75-18	0.210	0.233
	Lower	9	17	13-22		
Na (mmol/L)	Upper	20	138	136.75-140.25	0.324	0.183
	Lower	9	138	136-138		
CRP (mg/L)	Upper	20	0.95	0.60-1.93	0.383	0.162
	Lower	9	1.00	0.70-5.20		
NEU ($\times10^9$/L)	Upper	20	4.52	3.73-6.28	0.300	0.193
	Lower	9	5.58	4.31-6.45		
LYM ($\times10^9$/L)	Upper	20	2.47	2.04-2.71	0.310	0.193
	Lower	9	2.93	2.38-3.15		
MONO ($\times10^9$/L)	Upper	20	0.57	0.51-0.62	0.203	0.237
	Lower	9	0.65	0.57-0.93		
NLR	Upper	20	1.81	1.23-2.72	0.777	0.053
	Lower	9	2.16	1.23-3.08		
PLR	Upper	20	130.20	103.98-157.97	0.750	0.035
	Lower	9	119.45	83.49-146.21		
SII	Upper	20	529.55	344.01-871.98	0.509	0.123
	Lower	9	563.20	357.89-1078.67		
SIRI	Upper	20	1.09	0.73-1.66	0.437	0.144
	Lower	9	1.42	0.74-2.65		

*: Mann-Whitney U test, **: Effect size: $r = |Z|/\sqrt{N}$, GIB: Gastrointestinal bleeding, ALT: Alanine aminotransferase, Na: Sodium, CRP: C-reactive protein, NEU: Neutrophil, LYM: Lymphocyte, MONO: Monocyte, NLR: Neutrophil-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, SII: Systemic immune-inflammation index, SIRI: Systemic inflammation response index

Table 4. Significant correlations between length of hospital stay and laboratory parameters and inflammatory indices

Parameters	n	Spearman's rho (ρ)	p-value
Cr	29	-0.531*	0.003
ALB	29	-0.616*	<0.001
Ca	29	-0.516*	0.004
WBC	29	0.408*	0.028
HGB	29	-0.449*	0.015
HCT	29	-0.562*	0.002
NEU	29	0.511*	0.005
SIRI	29	0.450*	0.014

*: Correlation is significant at the 0.05 level (two-tailed), Cr: Creatinine, ALB: Albumin, Ca: Calcium, WBC: White blood cell, HGB: Hemoglobin, HCT: Hematocrit, NEU: Neutrophil, SIRI: Systemic inflammation response index

with identification of a bleeding source on upper GI endoscopy (13). Similarly, a multicenter study of 180 pediatric upper GI bleeding cases associated HGB <8 g/dL with significant lesions requiring endoscopic treatment (14). Nevertheless, substantial anemia may also occur in lower GI bleeding; a large pediatric cohort demonstrated lower HGB levels, particularly in bleeding associated with Meckel diverticulum (15). Thus, our nonsignificant findings suggest that anatomical bleeding location alone may not determine hematological loss; however, the limited sample size precludes definitive conclusions regarding differences between groups.

AST levels were also higher in the upper GI bleeding group, although the difference remained borderline (p=0.054). Portal hypertension and chronic liver disease may represent important etiologies of pediatric upper GI bleeding (16). However, our data did not allow linking AST elevation to a specific etiology. Therefore, this borderline finding should be interpreted cautiously. Overall, HGB, HCT, and AST alone may be insufficient to distinguish bleeding location in pediatric GI bleeding.

Biochemical parameters (ALT, Na, and CRP) and inflammatory markers (NEU, LYM, MONO, NLR, PLR, SII, and SIRI) did not differ significantly between upper and lower GI bleeding groups. The considerable etiological heterogeneity of pediatric GI bleeding may produce varying inflammatory responses regardless of anatomical location. In our lower GI bleeding group, etiologies were distributed as follows: colitis in 6 patients (66.7%), hemorrhoids in 2 patients (22.2%), and ileitis in 1 patient (11.1%). This etiological heterogeneity, encompassing conditions with potentially different degrees of inflammatory activity, may have contributed to the absence of significant differences in inflammatory indices between the upper- and lower-GI bleeding groups. Recent pediatric lower GI bleeding series have demonstrated diverse etiologies, ranging from inflammatory bowel disease to anal fissures and polyps (17). Similarly, pediatric upper GI bleeding encompasses gastric and duodenal lesions, infectious causes, and other etiologies (18). NLR, PLR, SII, and SIRI reflect different cellular components of systemic inflammation and may increase to varying degrees in pediatric inflammatory diseases (19,20). Accordingly, the lack of discriminatory value of these indices between bleeding groups may indicate that they reflect the inflammatory burden and clinical severity of the underlying disease rather than the anatomical location of bleeding.

Moderate negative correlations between hospital stay and HGB, HCT, ALB, Ca, and Cr, and positive correlations between hospital stay and WBC, NEU, and SIRI suggest that hematological, nutritional, and inflammatory status may be associated with the clinical course of pediatric GI bleeding. In particular, the association between lower HGB and HCT and longer hospitalization may indicate that greater blood loss warrants closer monitoring and treatment. Previous evidence showing associations of indices combining HGB and ALB with hospital stay and clinical outcomes in pediatric GI surgery supports these findings (21). The negative correlation between ALB and hospital stay may reflect its relationship with nutritional and inflammatory status. The inverse association with creatinine warrants caution because pediatric creatinine levels are strongly influenced by age and muscle mass, and may not reflect disease severity.

Positive associations of WBC and NEU with hospital stay suggest that systemic inflammation may accompany a prolonged clinical course. SIRI, which integrates NEU, MONO, and LYM, was also positively correlated with length of hospital stay. Pediatric studies support its association with inflammatory burden: higher SIRI was associated with complicated disease and longer hospitalization in necrotizing pneumonia (22), while children with pneumonia showed higher WBC, NEU, and SIRI than healthy controls (23). Elevated SIRI and SII were also associated with fulminant pediatric myocarditis (24). In adults with GI bleeding, SIRI was independently associated with in-hospital mortality, suggesting potential prognostic value (25). However, the performance of inflammatory indices may vary across pediatric populations (26). In our cohort, NLR, PLR, and SII were not significantly associated with hospital stay, whereas SIRI was significantly associated with hospital stay. This preliminary finding may suggest that SIRI captures a distinct aspect of systemic inflammation through its incorporation of the MONO component and that it warrants validation in larger cohorts.

Study Limitations

This study has several limitations. First, its retrospective, single-center design limits the generalizability of the findings to other pediatric populations. The relatively small sample size, particularly in the lower GI bleeding group, may have reduced statistical power to detect between-group differences, increasing the risk of a Type II error; therefore, non-significant findings should not be interpreted as definitive evidence of no difference between groups. Etiological heterogeneity may also have influenced laboratory and inflammatory parameters. Additionally, evaluating only admission laboratory values prevented the assessment of dynamic changes during the clinical course. The limited sample size precluded multivariable regression analysis to adjust for potential confounders, including age, bleeding location, underlying etiology, and treatment requirements or endoscopic interventions, particularly because pediatric parameters are age-dependent. Therefore, the observed associations require validation in larger pediatric cohorts.

Conclusion

This study evaluated associations of routine laboratory parameters and inflammatory indices with bleeding location and length of hospital stay in pediatric patients with GI bleeding. Most parameters, including NLR, PLR, SII, and SIRI, did not differ significantly between the upper and lower GI bleeding groups. However, given the limited sample size and the risk of Type II error, these findings should not be interpreted as definitive evidence of no difference. Hospital stay was negatively correlated with HGB, HCT, ALB, Ca, and Cr, and positively correlated with WBC, NEU, and SIRI. Notably, SIRI was significantly associated with length of hospital stay, whereas NLR, PLR, and SII were not, suggesting its potential complementary role in reflecting the systemic inflammatory burden. However, SIRI and other laboratory parameters should not be considered standalone prognostic markers and should be interpreted alongside clinical findings, hemodynamic status, bleeding severity, and endoscopic evaluation. These preliminary associations require validation in larger, multicenter prospective pediatric studies before their prognostic significance and clinical utility can be established.

Ethics

Ethics Committee Approval: Ethical approval was obtained from the Malatya Turgut Özal University Health Sciences Scientific Research Ethics Committee (approval date: 11.09.2026; approval no: E-30785963-020-421429).

Informed Consent: Because the study was designed retrospectively, no written informed consent form was obtained from the patients.

Generative AI Usage Statement: The authors used artificial intelligence-based language editing tools solely to improve grammar and linguistic clarity during manuscript preparation.

Footnotes

Authorship Contributions

Concept: S.N., N.A.M., Design: S.N., Data Collection or Processing: S.N., Analysis or Interpretation: S.N., N.A.M., Literature Search: S.N., Writing: S.N., N.A.M.

Conflict of Interest: The authors declare that they have no conflicts of interest.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Piccirillo M, Pucinischi V, Mennini M, Strisciuglio C, Iannicelli E, Giallorenzi MA, et al. Gastrointestinal bleeding in children: diagnostic approach. Ital J Pediatr. 2024;50:13.
2. Novak I, Bass LM. Gastrointestinal bleeding in children: current management, controversies, and advances. Gastrointest Endosc Clin N Am. 2023;33:401-21.
3. Kostyuk O, Luyt K. Pediatric gastrointestinal bleeding: identification and management in the emergency department. Pediatr Emerg Med Pract. 2024;21:1-32.
4. Sur LM, Armat I, Sur G, Tisa IB, Bordea MA, Lupan I, et al. Practical aspects of upper gastrointestinal bleeding in children. J Clin Med. 2023;12:2921.
5. Bello FPS, Cardoso S, Tannuri AC, Preto-Zamperlini M, Schvartsman C, Farhat SCL. Acute upper gastrointestinal bleeding due to portal hypertension in children: what is the best timing of endoscopy? Dig Liver Dis. 2022;54:63-8.
6. Yucel A. Evaluation of the Sheffield score, clinical characteristics, and the therapeutic approach in children with upper gastrointestinal system bleeding. Eur Rev Med Pharmacol Sci. 2022;26:7986-95.
7. Yabe K, Kouchi K, Takenouchi A, Matsuoka A, Kudou W, Nakata C. Current status and future challenges in the endoscopic management of non-variceal upper gastrointestinal bleeding in children. Pediatr Surg Int. 2020;36:1451-8.
8. Schluckebier D, Afzal NA, Thomson M. Therapeutic upper gastrointestinal endoscopy in pediatric gastroenterology. Front Pediatr. 2022;9:715912.
9. Caltekin I, Demir B, Atik D, Albayrak L, Gokcen E, Savrun A, et al. Evaluation of red blood cell distribution width and mean platelet volume levels in hypertensive and non-hypertensive epistaxis patients. Ann Med Res. 2020;27:938-42.
10. Cumaoglu MO, Vural A, Alpaslan M, Dolanbay T, Kulu AR, Sezgin E, et al. Inflammatory indices as independent predictors of in-hospital mortality in acute upper gastrointestinal bleeding. BMC Gastroenterol. 2026;26:127.
11. Zhang X, Song Y, Yin M, Zhang L. Establishment of age-specific reference intervals for peripheral blood SII, NLR, PLR, and LMR in healthy children. BMC Immunol. 2026;27:22.
12. Xu QC, Liu ZY, Gao F, He YT, He XR, Zheng W, et al. Advances in standardized diagnosis and management of pediatric gastrointestinal bleeding. World J Pediatr. 2025;21:1218-29.
13. Iwama I, Yoshida M, Hara T, Nambu R. Causes of melena and effective examination strategies in children. Front Pediatr. 2021;9:780356.
14. Acharyya BC, Mukhopadhyay M, Chakraborty H. Changing trend in the spectrum of upper gastrointestinal bleeding in children-A multicentre experience. Indian J Gastroenterol. 2023;42:562-8.

15. Couselo M, Ibáñez V, Pemartín B, Fonseca R, Vila JJ. Diagnostic value of hemoglobin concentration in identifying Meckel's diverticulum in pediatric lower gastrointestinal bleeding. Pediatr Surg Int. 2024;41:50.
16. Kocic M, Rasic P, Marusic V, Prokic D, Savic D, Milickovic M, et al. Age-specific causes of upper gastrointestinal bleeding in children. World J Gastroenterol. 2023;29:6095-110.
17. Bircan KM, Elmas A, Akçam M. Pediatric lower gastrointestinal bleeding: endoscopic and histopathologic spectrum in a 16-year tertiary-center cohort. Biomedicines. 2026;14:1645.
18. Galos F, Ionescu MI, Mirea MDL, Boboc AA, Ioan A, Boboc C. Trends in upper gastrointestinal bleeding in children: the impact of *Helicobacter pylori* infection and non-steroidal anti-inflammatory drug use. Antibiotics (Basel). 2024;13:752.
19. Siki FÖ, Sarıkaya M, Gunduz M, Sekmenli T, Korez MK, Ciftci I. Evaluation of the systemic immune inflammation index and the systemic inflammatory response index as new markers for the diagnosis of acute appendicitis in children. Ann Saudi Med. 2023;43:329-38.
20. Elmas A, Akçam M. Diagnostic performance of blood-based inflammatory indices, including a novel composite score, for estimating fecal calprotectin levels in pediatric inflammatory bowel disease. J Clin Med. 2026;15:3046.
21. Silahli M, Tekin M. Albumin haemoglobin index: a novel pre-operative marker for predicting mortality and hospital stay in patients under one-year undergoing gastrointestinal surgeries. Afr J Paediatr Surg. 2022;19:89-96.
22. Elmeazawy R, Ayoub D, Morad LM, El-Moazen AMF. Role of systemic immune-inflammatory index and systemic inflammatory response index in predicting the diagnosis of necrotizing pneumonia in children. BMC Pediatr. 2024;24:496.
23. Yucel K, Disci SI. Systemic immune inflammation index, systemic inflammation response index, aggregate index of systemic inflammation, and follistatin-like protein-1 levels in children diagnosed with pneumonia. Pediatr Allergy Immunol Pulmonol. 2024;37:106-11.
24. Kangel D, Ozyılmaz İ, Ozkok S, Özcanoğlu HD, Güzelbağ AN, Çevlik B, et al. New systemic inflammatory indices as predictors of fulminant myocarditis in children. Diagnostics (Basel). 2025;15:961.
25. Erdoğan Ç, İnan B, Ateş İ, Kılıç ZMY. Systemic inflammation response index (SIRI) and aggregate index of systemic inflammation (AISI) as predictors of mortality in patients with upper gastrointestinal bleeding. J Clin Med. 2026;15:2245.
26. Wang S, Wan Y, Zhang W. The clinical value of systemic immune inflammation index (SII) in predicting the severity of hospitalized children with mycoplasma pneumoniae pneumonia: a retrospective study. Int J Gen Med. 2024;17:935-42.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.63139
Eurasian J Emerg Med. 2026;25: 481-8

Changes in Serum Mineral and Trace Element Levels Following Laparoscopic Cholecystectomy: A Prospective Pilot Study

Serpil Erşan[1], Hacı Bolat[2]

[1]Niğde Ömer Halisdemir University Faculty of Medicine, Department of Medical Biochemistry, Niğde, Türkiye
[2]Niğde Private Hayat Hospital, Clinic of General Surgery, Niğde, Türkiye

Abstract

Aim: The gallbladder is increasingly recognized as a metabolically active organ that regulates bile acid signaling and systemic metabolic homeostasis. While the metabolic consequences of laparoscopic cholecystectomy (LC) have been extensively studied, its long-term impact on mineral and trace element homeostasis remains unknown. To our knowledge, this is the first prospective study to evaluate the longitudinal changes in multiple serum minerals and trace elements following LC. Therefore, this study aimed to assess changes in serum mineral and trace element levels and body mass index (BMI) at six months after LC in patients with symptomatic gallstone disease.

Materials and Methods: This prospective longitudinal pilot study included 30 patients with symptomatic gallstone disease who underwent LC. Serum samples were collected preoperatively and at six months postoperatively. Serum levels of Cu, Zn, Fe, Mg, Cr, Cd, and Se were measured using inductively coupled plasma mass spectrometry. BMI was calculated by measuring weight before surgery and six months after surgery. Pre- and postoperative measurements and sex-based differences were statistically analyzed, with correction for multiple comparisons by sex.

Results: BMI increased significantly six months after LC, compared with baseline ($p<0.001$). Serum Fe, Zn, Mg, and Cr levels decreased significantly after LC ($p<0.001$), while Cu and Cd levels increased significantly ($p<0.001$). Although a slight decrease in Se levels was observed, it was not statistically significant ($p=0.856$). After correction for multiple comparisons, women had significantly lower preoperative Cr levels and postoperative Fe levels than men.

Conclusion: LC was associated with significant changes in serum mineral and trace element homeostasis and an increase in BMI six months postoperatively. These findings may be related to alterations in BA physiology and enterohepatic circulation following LC; however, the underlying mechanisms remain to be elucidated. Long-term prospective studies integrating BA profiles, gut microbiota, and nutritional biomarkers are required to elucidate the underlying mechanisms of these changes.

Keywords: Laparoscopic cholecystectomy, trace elements, iron, zinc, magnesium, chromium, copper, cadmium, body mass index, bile acids

Introduction

Gallstone disease (GSD) is one of the most common surgical diseases worldwide and, with its increasing incidence over time, represents an important health problem (1). Cholelithiasis, defined as the formation of concentrated bile crystals or stones within the gallbladder (GB), is the most common GB disease. Cholelithiasis is characterized by the formation of crystallized deposits, usually composed of cholesterol or bilirubin, in the GB or bile ducts. Its global prevalence is approximately 6%, and it has been reported to be associated with geographic origin, age, sex, and metabolic disorders (2). Although patients are generally asymptomatic, invasive treatment modalities, predominantly surgical approaches, are used for patients who are symptomatic (1). Today, laparoscopic cholecystectomy (LC) is accepted as the gold-standard surgical method in the treatment of symptomatic GSD because, owing to its minimally invasive nature, it provides faster recovery, shorter hospital stay, and lower postoperative morbidity compared with open surgery (1,3). Although cholecystectomy has been reported to be a safe procedure in the treatment of patients with gallstones, cholecystectomy has also been associated with metabolic disorders, including type

Corresponding Author: Serpil Erşan MD, Niğde Ömer Halisdemir University Faculty of Medicine, Department of Medical Biochemistry, Niğde, Türkiye
E-mail: serpilersan@gmail.com **ORCID ID:** orcid.org/0000-0002-8968-4044

Cite this article as: Erşan S, Bolat H. Changes in serum mineral and trace element levels following laparoscopic cholecystectomy: a prospective pilot study. Eurasian J Emerg Med. 2026;25: 481-8

Received: 25.07.2026
Accepted: 02.10.2026
Published: 26.10.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

2 diabetes mellitus and metabolic dysfunction-associated fatty liver disease. While the GB has traditionally been regarded as a reservoir organ that stores bile, recent evidence has shown that it also has important direct and indirect roles in metabolic regulation (3,4). The primary function of the GB is to store and concentrate bile produced by the liver (5). However, in addition to the well-known roles of bile within the GB in the digestion and absorption of lipids, it also has physiological roles in glucose, lipid, and energy homeostasis (4). Evidence obtained from studies has shown that the elevated cholesterol levels and impaired insulin resistance observed after cholecystectomy indicate that the GB plays a role in metabolic regulation (4,6). Bile acids (BAs) also act as signaling molecules that regulate fundamental metabolic pathways via nuclear receptors such as the farnesoid X receptor (FXR) (6,7). Therefore, unexplained metabolic abnormalities in patients who have undergone cholecystectomy may be related to the different physiological functions of the GB (8,9). As a result of cholecystectomy-related loss of the GB's reservoir and concentrating capacity, various functions, including the regulation of the enterohepatic circulation of concentrated BAs, are altered (10). In this situation, the diluted bile continuously produced by the liver is secreted into the intestine (9). The altered secretion of BAs after cholecystectomy has adverse effects on intestinal functions (11,12).

Minerals and trace elements comprise a heterogeneous group with diverse biological and toxicological roles. While elements such as Fe, Zn, Mg, Cu, and Se contribute to essential physiological processes, Cd is a non-essential toxic element, and the essentiality of Cr in humans remains debated (13). Studies have shown that conjugated bile salts have superior emulsification properties compared with unconjugated ones and facilitate the absorption of lipophilic compounds, including fat-soluble vitamins (14). However, it is thought that changes in BA physiology may alter the intestinal bioavailability and absorption of certain trace elements by affecting micelle stability and the environment of the intestinal lumen (15,16).

Although there is growing evidence that cholecystectomy may affect metabolic homeostasis through changes in BA physiology, its long-term effects on serum mineral and trace element homeostasis have not yet been fully elucidated. Accordingly, the present pilot study included both nutritionally relevant elements (Fe, Zn, Mg, Cu, and Se) and elements of toxicological relevance or of uncertain biological relevance (Cd and Cr) to explore whether LC is associated with broader changes in the serum elemental profile. In this study, we aimed to evaluate changes in preoperative and 6-month postoperative serum levels of Cu, Zn, Fe, Mg, Cr, Cd, and Se in patients who underwent LC.

Materials and Methods

Study Groups

This study included 30 patients with symptomatic cholelithiasis who underwent LC at the General Surgery Clinic of Private Niğde Hayat Hospital. A diagnosis of symptomatic cholelithiasis was made by abdominal ultrasonography in patients presenting with symptoms such as postprandial abdominal pain, abdominal distension, and pain radiating to the right shoulder (17). Patients with acute cholecystitis, choledocholithiasis, diabetes, hypertension, nephrotic syndrome, pancreatitis, anaemia, familial hyperlipidaemia, hyperthyroidism or hypothyroidism, coronary artery disease, other metabolic diseases, morbid obesity (body mass index, BMI ≥40, pregnancy, those using lipid-lowering therapy, those with hormonal disorders, and patients who had undergone any small intestine surgery were not included in the study. The study was approved by the Non-Interventional Clinical Research Ethics Committee of Niğde Ömer Halisdemir University (approval no: 2025/70, date: 29.07.2025), and patients who voluntarily agreed to participate were enrolled in the study. Written informed consent was obtained from all participants prior to enrollment in the study. All study procedures were carried out in accordance with the ethical principles of the Declaration of Helsinki.

Collection of Blood Samples

For each participant, 5 mL of blood was drawn into two sterile tubes containing gel. Similarly, when the patients attended the general surgery clinic for follow-up, blood samples were collected 6 months after LC. Based on data from previous studies, it has been shown that lipid levels are affected by the time elapsed after cholecystectomy, with the most pronounced changes observed at six months after surgery (18). Therefore, therefore, in our study, it was preferred to collect blood samples 6 months after LC. Blood samples collected in gel-containing tubes before and after LC were allowed to clot at room temperature, and then the clotted samples were centrifuged at 3000 × g for 10 minutes to separate the serum phase. Serum samples were stored at -80 °C until mineral and trace-element levels were analyzed.

Determination of BMI

The body weights and heights of participants in the study group were measured before and after LC surgery. BMI values were calculated by dividing body weight in kilograms by the square of height in metres (kg/m^2). According to the World Health Organization's diagnostic criteria for obesity, individuals with a BMI ≥30 kg/m^2 are considered obese (19).

Measurement of Mineral and Trace Element Levels

Serum mineral and trace element levels (Cu, Cr, Cd, Mg, Fe, Zn, and Se) were determined using inductively coupled plasma mass spectrometry (ICP-MS; Thermo Scientific iCAP RQ, Bremen, Germany), which is widely accepted as the reference analytical technique for multi-element determination in biological samples because of its excellent analytical accuracy, a wide linear dynamic range, and the ability to measure multiple elements simultaneously at trace levels. Prior to ICP-MS analysis, samples were subjected to microwave-assisted digestion (MARS 6, CEM Corporation, Matthews, NC, USA). Calibration curves were constructed using multi-element standard solutions with terbium (Tb) as the internal standard. The calibration curves showed good linearity (R^2 >0.99) for all elements. The limits of detection for the analysed elements were in the ppb range and were determined from calibration curves generated by the ICP-MS system. Throughout the analysis, quality control measures such as blanks and standard evaluations were applied. Each sample was analysed in triplicate and findings were recorded in ppb or µg/L.

Statistical Analysis

Numerical variables were summarized as mean ± standard deviation (SD) and minimum-maximum values. Normality was assessed using the Shapiro-Wilk test and Q-Q plots. Preoperative and postoperative measurements were compared using paired-samples t-tests or Wilcoxon signed-rank tests, as appropriate. Effect sizes were reported as Cohen's dz for paired t-tests and matched-pairs rank-biserial correlation (rrb) for Wilcoxon tests. Sex-based comparisons were performed using Welch's independent-samples t-test, with Hedges' g reported as the effect size. Bonferroni correction was applied to the 16 sex-based comparisons, yielding an adjusted significance threshold of p<0.003. Statistical analyses were performed using IBM SPSS version 26 (SPSS, Inc., Chicago, IL, USA). A two-sided p-value <0.05 was considered statistically significant, except for Bonferroni-adjusted comparisons.

Results

The mean age of the patients with symptomatic gallstones included in the study was 44.73 years (range 31-67). Of these, 63.33% were female (n=19) and 36.66% were male (n=11). A statistically significant increase was observed in BMI from 28.4±3.10 kg/m^2 before LC to 29.8±3.97 kg/m^2 six months after LC (Table 1; p<0.001). The proportion of patients with BMI ≥ 30 kg/m^2 increased from 33.3% (10/30) before LC to 46.7% (14/30) six months after LC. Mineral and trace element levels in serum samples collected from patients before LC and 6 months after LC were compared. Cr, Fe, Zn, and Mg levels decreased after LC (Table 1; p<0.001), whereas Cu and Cd levels increased (Table 1; p<0.001). Although serum Se levels were lower after LC compared with pre-LC values, this decrease was not statistically significant (Table 1; p=0.856). After Bonferroni correction for multiple comparisons (adjusted significance threshold, p<0.003), women had significantly lower preoperative Cr levels and lower postoperative Fe levels than men. No other sex-based comparisons met the Bonferroni-adjusted significance threshold (Table 2).

Discussion

Increasing evidence suggests that the GB is not merely a passive bile reservoir but an active regulator of systemic metabolic homeostasis (6,7,20-25). While the metabolic consequences of cholecystectomy on glucose and lipid metabolism and on fatty liver disease have been extensively investigated, its potential impact on mineral and trace element homeostasis remains almost entirely unexplored. To our knowledge, this is the first

Table 1. Comparison of BMI and mineral and trace element levels before and after LC operation

	Before LC Mean ± SD (min-max)	After LC Mean ± SD (min-max)	p	Effect size
BMI	28.4±3.10 (22.8-34.4)	29.8±3.97 (22.5-39.8)	r_rb	r_rb
Cr	4.09±0.691 (3.01-5.55)	3.62±0.640 (2.83-5.18)	**<0.001**[a]	r_rb=-1.00
Cu	848±221 (506-1361)	1270±458 (591-2200)	**<0.001**[a]	r_rb=1.00
Fe	1178±130 (836-1372)	970±127 (757-1257)	**<0.001**[b]	dz=-1.48
Zn	691±127 (404-892)	574±113 (387-796)	**<0.001**[b]	dz=-1.01
Mg	7920±853 (6371-9734)	6610±694 (5310-7960)	**<0.001**[b]	dz=-2.12
Se	66.0±5.32 (54.4-77.4)	65.9±5.54 (53.6-76.3)	0.856[b]	dz=-0.03
Cd	31.1±2.16 (23.1-36.5)	38.0±3.40 (29.5-43.9)	**<0.001**[b]	dz=2.45

[a]Wilcoxon signed-rank test, [b]paired-samples t-test. Effect sizes are reported as matched-pairs rank-biserial correlation (r_rb) for Wilcoxon signed-rank tests and Cohen's dz for paired-samples t-tests
BMI: Body mass index, SD: Standard deviation, Min-max: Minimum-maximum

prospective study to demonstrate that LC is associated with persistent changes in several circulating minerals and trace elements six months postoperatively.

It is increasingly accepted that changes in intestinal physiology, nutrient absorption, and enterohepatic circulation following LC may have long-term effects on metabolic homeostasis. Given that these changes may also affect the intestinal absorption and metabolism of trace elements, this study evaluated changes in serum trace element levels before LC and six months postoperatively. Among the symptomatic patients with gallstones who participated in the study, 63.33% were female and 36.66% were male; the average age was 44.7 years. The higher number of female patients in our study group is consistent with the higher prevalence of GSD in women reported in the literature (2,26-28).

Although trace elements constitute only a small part of the diet, they play important roles in physiological and biochemical processes (29). Adequate intake of trace elements is essential for maintaining human metabolic homeostasis, and imbalances can lead to the development of cardiovascular, digestive, endocrine, blood, and other diseases (20). Although LC is considered a safe and effective treatment method for symptomatic GSD, there is increasing evidence that changes in intestinal physiology and nutrient absorption following surgery can affect metabolic homeostasis (3,7). The effects of these changes on serum trace element homeostasis remain unclear.

In our study, six months after LC, serum Fe, Zn, Cr, and Mg levels decreased, whereas Cd and Cu levels increased (all $p<0.001$). Fe and Zn are absorbed primarily in the proximal small intestine, and their intestinal bioavailability is influenced by luminal conditions and interactions with other minerals (30,31). Therefore, maintaining their solubility in the lumen is critical for efficient absorption. Bile salts play an important role in the physicochemical environment of the intestinal lumen and facilitate the micellar solubilization and absorption of lipids and fat-soluble compounds (14,32).

Cholecystectomy alters the pattern of BA delivery to the intestine and may modify enterohepatic circulation and the intestinal microenvironment (33,34). In this context, the decreases in serum Fe and Zn observed in our study may partly reflect changes in their intestinal bioavailability associated with altered bile flow and changes in the intestinal environment after cholecystectomy. However, as intestinal absorption and BA profiles were not directly assessed in the present study, this proposed mechanism remains speculative.

Naik et al. (35) reported a high prevalence of iron deficiency among symptomatic patients with gallstones, suggesting that iron metabolism may be closely related to GSD. However, this study only evaluated the preoperative period and did not examine changes in Fe homeostasis after cholecystectomy. In our study, significantly lower serum Fe levels observed six months after LC

Table 2. Comparison of demographic characteristics and serum mineral and trace element levels according to sex

		Female (n=19) Mean ± SD (Min-max)	Male (n=11) Mean ± SD (Min-max)	p	Hedges' g
Age (year)		44.16±9.82 (31.00-67.00)	45.73±7.71 (37.00-61.00)	0.632	-0.17
BMI		29.20±3.32 (23.53-34.38)	27.13±2.25 (22.84-30.42)	0.053	0.67
Cr	Before LC	3.80±0.60 (3.01-5.12)	4.58±0.56 (3.70-5.55)	**0.001***	-1.31
	After LC	3.41±0.50 (2.83-4.69)	3.99±0.71 (2.83-5.18)	0.028	−0.98
Cu	Before LC	884,19±242,97 (506,20-1361,21)	784,42±168,06 (525,02-1029,39)	0.196	0.44
	After LC	1300,25±505,66 (591,19-2157,04)	1230,30±381,34 (822,38-2200,20)	0.658	0.15
Fe	Before LC	1148,80±134,88 (836,05-1336,60)	1232,63±105,27 (1077,14-1372,02)	0.064	-0.67
	After LC	914,35±95,27 (756,63-1117,68)	1067,08±119,28 (878,10-1257,05)	**0.002***	-1.42
Zn	Before LC	679,45±124,71 (403,80-891,60)	709,71±135,88 (472,90-891,60)	0.552	-0.23
	After LC	569,07±110,95 (407,10-796,00)	581,37±119,91 (387,30-782,60)	0.786	-0.10
Mg	Before LC	7773,25±801,13 (6371,40-9206,50)	8172,48±918,21 (6372,20-9734,00)	0.245	-0.46
	After LC	6470,24±574,92 (5309,90-7825,90)	6861,29±833,58 (5704,80-7960,30)	0.188	-0.56
Se	Before LC	66,54±5,53 (59.72-77.40)	65.19±5.06 (54.38-70.85)	0.505	0.24
	After LC	66.17±5.65 (54.37-76.30)	65.31±5.55 (53.64-73.03)	0.687	0.15
Cd	Before LC	30.83±2.17 (23.09-33.96)	31.63±2.15 (29.31-36.45)	0.336	-0.36
	After LC	37.74±3.48 (29.51-43.96)	38.36±3.38 (32.13-43.51)	0.634	-0.18

Data are presented as mean ± SD. [a]Welch's independent-samples t-test. Effect sizes are reported as Hedges' g. Bonferroni correction was applied for 16 comparisons, resulting in an adjusted significance threshold of $p<0.003(0.05/16)$. LC: Laparoscopic cholecystectomy, BMI: Body mass index; SD: standard deviation

suggest that alterations in Fe status may persist after surgery or be associated with postoperative changes in biliary physiology. However, given the prospective pre-post observational design of our study, a causal relationship cannot be established, and this hypothesis requires confirmation in larger, long-term prospective studies. A notable decrease in serum Fe concentrations and a significant increase in serum Cd concentrations were observed six months after LC. Fe and Cd can share intestinal metal transport pathways, and experimental studies suggest that alterations in Fe status may influence intestinal Cd uptake (36-38). Consequently, the decrease in serum Fe observed in our study may have contributed to the increase in Cd concentrations through shared metal transport pathways. However, because metal transporter expression and intestinal Cd absorption were not directly assessed, this interpretation should be regarded as a plausible mechanistic hypothesis rather than evidence of a causal relationship. Furthermore, serum Cd concentrations may be influenced by environmental exposure, dietary factors, and other physiological factors that were not comprehensively assessed in the present study (38).

Magnesium absorption in the intestine occurs through both passive paracellular diffusion and active transcellular transport involving the transient receptor potential melastatin channels TRPM6 and TRPM7 (39). Magnesium absorption efficiency is influenced by luminal factors, including pH, mineral solubility, food matrix composition, and intestinal transit time (39,40). In our study, serum Mg levels decreased significantly six months after LC. Changes in BA dynamics and the intestinal environment following cholecystectomy may influence intestinal mineral absorption and could therefore contribute to the observed change in serum Mg (34). However, intestinal Mg absorption, BA profiles, and related transport mechanisms were not directly assessed in the present study; therefore, this interpretation remains hypothesis-generating. Additionally, in patients experiencing BA diarrhea post-cholecystectomy, accelerated intestinal transit and increased gastrointestinal losses may further diminish magnesium absorption (41).

Supporting our findings, a previous study noted reduced serum and erythrocyte magnesium concentrations in patients with post-cholecystectomy syndrome and reported symptomatic improvement with magnesium supplementation. However, the lack of a control group and the selective patient population limit the causal interpretation of the association between magnesium status and post-cholecystectomy syndrome (42). Interactions between divalent minerals, fatty acids, and other lipophilic components during intestinal digestion may influence their physicochemical behavior and bioavailability (16). Such nutrient-mineral interactions could represent a mechanism contributing to the postoperative change in serum Mg observed in our study. However, because fat digestion, intestinal free fatty acid concentrations, and Mg absorption were not directly assessed, this possibility remains speculative. Rather than indicating that cholecystectomy directly causes magnesium deficiency, our findings suggest that altered BA dynamics, changes in the intestinal environment, and potential nutrient-mineral interactions may collectively contribute to the observed postoperative decrease in serum Mg.

In our study, serum Cu levels increased significantly six months after LC compared with preoperative levels. Copper homeostasis is regulated through coordinated intestinal absorption. Fraser et al. (43) reported early postoperative changes in serum Fe and Zn concentrations following cholecystectomy and cardiac surgery, whereas serum Cu remained relatively stable during the early postoperative period. These findings highlight the influence of the acute surgical response on circulating trace element concentrations. In contrast, our study assessed serum Cu levels over a longer period by comparing preoperative values with those obtained six months after surgery. Therefore, the observed increase in serum Cu is less likely to reflect solely a transient acute-phase response to surgical trauma. Postoperative changes in hepatobiliary physiology and intestinal homeostasis may contribute to the observed increase in serum Cu; however, because copper absorption, biliary excretion, and related regulatory mechanisms were not directly assessed, this interpretation remains hypothesis-generating.

Six months after LC, serum Se levels did not change significantly compared with preoperative levels, whereas serum Cr levels decreased significantly. These findings suggest that the postoperative changes observed in our study were not uniform across all trace elements, potentially reflecting differences in their absorption, transport, and homeostatic regulation. Selenium is generally well absorbed, although its bioavailability varies according to its chemical form, dietary composition, and interactions with the gut microbiota (44-46). These characteristics may partly explain the relative stability of serum Se levels observed in our study. In contrast, the intestinal absorption and bioavailability of Cr are relatively low and may be influenced by dietary and gastrointestinal factors (47). Therefore, changes in the intestinal environment following cholecystectomy could contribute to the observed decrease in serum Cr. Because dietary intake, intestinal absorption, gut microbiota, and element-specific transport mechanisms were not directly assessed, these interpretations are hypothesis-generating and do not provide evidence of a causal relationship.

In this study, BMI values recorded prior to surgery and six months after LC showed a significant postoperative increase.

This finding aligns with previous research, which has also documented an increase in BMI following cholecystectomy (21,22,33,48). In a cross-sectional study, Cao et al. (21) reported that cholecystectomy was independently associated with higher BMI, with a more pronounced association among younger adults. These findings suggest that the metabolic changes associated with cholecystectomy may extend beyond its effects on bile storage and digestion, potentially involving broader alterations in metabolic homeostasis. Although postoperative dietary changes may also contribute to weight gain, dietary intake was not monitored in our study; therefore, the relative contribution of dietary and metabolic factors to the observed increase in BMI could not be determined.

Recent research suggests that the GB functions not only as a passive bile reservoir but also as an active participant in the regulation of systemic metabolic homeostasis (6,7,20,25). Following cholecystectomy, the physiological meal-dependent pattern of BA delivery to the intestine becomes less pronounced, with bile entering the duodenum in a more continuous manner. This change may alter enterohepatic circulation and the BA pool composition, thereby affecting metabolic signaling pathways mediated by the FXR, Takeda G protein-coupled receptor 5, and fibroblast growth factor 19. Given that these pathways are involved in the regulation of glucose metabolism, energy expenditure, lipid homeostasis, and insulin sensitivity, alterations in BA signaling have been proposed as a potential mechanism contributing to long-term weight gain and metabolic disturbances following cholecystectomy (6,7,21,22,24,34,49,50). These physiological changes may affect both systemic energy metabolism and intestinal absorption processes. Consequently, the simultaneous increase in BMI and changes in serum mineral and trace element levels observed in our study may reflect overlapping mechanisms involving post-cholecystectomy remodeling of the BA-gut microbiota-metabolic axis. Assessing changes in BMI alongside mineral and trace element levels is therefore an important aspect of this study. Although increased body weight is known to affect mineral and trace element metabolism, the distinct directional changes in individual analytes suggest that factors beyond weight change may also be involved; this possibility, however, could not be directly tested in the present study. Specifically, post-cholecystectomy systemic metabolic remodeling and element-specific intestinal transport mechanisms may contribute to the observed changes in mineral and trace element levels. However, given the limited number of studies on this topic, the proposed mechanisms should be considered hypothesis-generating rather than as evidence of a causal relationship.

Our findings expand the current understanding of cholecystectomy beyond its effects on BA physiology and lipid metabolism. The persistent changes in multiple serum mineral and trace element levels observed six months after cholecystectomy are consistent with broader metabolic remodeling. These results further support the emerging view that the GB serves not only as a passive reservoir for bile but also as an integral component of systemic metabolic regulation. Future large-scale mechanistic studies integrating BA profiling, gut microbiota characterization, intestinal mineral transport pathways, and metabolic biomarkers are needed to elucidate the biological mechanisms underlying these observations and to assess their potential clinical implications.

Study Limitations

This pilot study represents, to our knowledge, the first prospective longitudinal investigation of changes in serum mineral and trace element levels following LC. However, several limitations should be considered. First, the study included only 30 participants, which limits the precision and generalizability of the findings, particularly for sex-based subgroup comparisons. Second, the study used a pre-post observational design without a healthy or non-surgical control group. Therefore, the preoperative measurements should not be considered healthy baseline values because all participants had symptomatic GSD, and the observed longitudinal changes cannot be attributed solely to LC. In addition, disease-related gastrointestinal symptoms, dietary changes, medication use, and perioperative supportive interventions were not systematically quantified; therefore, their potential influence on preoperative serum mineral and trace element concentrations could not be evaluated. Furthermore, method-specific clinical reference intervals were not established in the study population. Therefore, the absolute serum concentrations should not be interpreted as evidence of element deficiency or toxicity. The present findings primarily reflect longitudinal within-patient changes. Third, individual BA profiles, gut microbiota, and relevant nutritional biomarkers were not assessed, limiting our ability to directly examine the proposed mechanisms linking postoperative changes in and intestinal physiology to mineral and trace element homeostasis. Furthermore, serum mineral and trace element levels were assessed at only two time points, immediately before LC and six months after surgery; therefore, early and intermediate postoperative changes could not be characterized. Urinary excretion of minerals and trace elements was also not assessed. An additional limitation is the lack of dietary monitoring throughout the study. Dietary assessment would have facilitated evaluation of the potential effects of habitual dietary intake

and postoperative dietary changes on serum mineral and trace element levels and BMI.

Conclusion

In this prospective pilot study, LC was associated with a significant increase in BMI and a distinct shift in the serum elemental profile six months after surgery, characterized by decreased Fe, Zn, Mg, and Cr levels and increased Cu and Cd levels, while Se levels remained unchanged. These findings suggest that the gallbladder may play a role in maintaining mineral and trace element homeostasis beyond its classical function in bile storage. Monitoring body weight and selected mineral and trace element levels may therefore be considered during the postoperative follow-up of patients undergoing LC, pending confirmation of these preliminary findings in larger controlled studies.

Ethics

Ethics Committee Approval: The study was approved by the Non-Interventional Clinical Research Ethics Committee of Niğde Ömer Halisdemir University (approval no: 2025/70, date: 29.07.2025), and patients who voluntarily agreed to participate were enrolled in the study.

Informed Consent: Written informed consent was obtained from all participants prior to enrollment in the study. All study procedures were carried out in accordance with the ethical principles of the Declaration of Helsinki.

Footnotes

Authorship Contributions

Surgical and Medical Practices: H.B., Concept: S.E., Design: S.E., Data Collection or Processing: H.B., S.E., Analysis or Interpretation: H.B., S.E., Literature Search: H.B., S.E., Writing: H.B., S.E.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Wang K, Liu Z, Tang R, Sha Y, Wang Z, Chen Y, et al. Gallstones in the era of metabolic syndrome: pathophysiology, risk prediction, and management. Cureus. 2025;17:e80541.
2. Wang X, Yu W, Jiang G, Li H, Li S, Xie L, et al. Global epidemiology of gallstones in the 21st century: a systematic review and meta-analysis. Clin Gastroenterol Hepatol. 2024;22:1586-95.
3. Zhang Y, Dai X, Duan R, Wei L. Research progress in the treatment of gallstones with laparoscopic and endoscopic surgery: a narrative review. BMC Surg. 2025;25:238.
4. Arrese M, Cortés V, Barrera F, Nervi F. Nonalcoholic fatty liver disease, cholesterol gallstones, and cholecystectomy: new insights on a complex relationship. Curr Opin Gastroenterol. 2018;34:90-6.
5. Jones MW, Weir CB, Marietta M. Gallstones (Cholelithiasis). 2026 Jun 19. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026.
6. Cortés VA, Barrera F, Nervi F. Pathophysiological connections between gallstone disease, insulin resistance, and obesity. Obes Rev. 2020;21:e12983.
7. Ciaula AD, Garruti G, Wang DQH, Portincasa P. Cholecystectomy and risk of metabolic syndrome. Eur J Intern Med. 2018;53:3-11.
8. Chavez-Tapia NC, Kinney-Novelo IM, Sifuentes-Rentería SE, Torres-Zavala M, Castro-Gastelum G, Sánchez-Lara K, et al. Association between cholecystectomy for gallstone disease and risk factors for cardiovascular disease. Ann Hepatol. 2012;11:85-9.
9. Sonne DP, Hare KJ, Martens P, Rehfeld JF, Holst JJ, Vilsbøll T, et al. Postprandial gut hormone responses and glucose metabolism in cholecystectomized patients. Am J Physiol-Gastrointest Liver Physiol. 2013;304:G413-9.
10. Yun S, Choi D, Lee KG, Kim HJ, Kang BK, Kim H, et al. Cholecystectomy causes ultrasound evidence of increased hepatic steatosis. World J Surg. 2016;40:1412-21.
11. Fort JM, Azpiroz F, Casellas F, Andreu J, Malagelada JR. Bowel habit after cholecystectomy: physiological changes and clinical implications. Gastroenterology. 1996;111:617-22.
12. O'Donnell LJ. Post-cholecystectomy diarrhoea: a running commentary. Gut. 1999;45:796-7.
13. Rolić T, Yazdani M, Mandić S, Distante S. Iron metabolism, calcium, magnesium and trace elements: a review. Biol Trace Elem Res. 2025;203:2216-25.
14. Moghimipour E, Ameri A, Handali S. Absorption-enhancing effects of bile salts. Molecules. 2015;20:14451-73.
15. Zuvarox T, Goosenberg E, Belletieri C. Malabsorption syndromes. 2025 Jul 7. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025.
16. Corte-Real J, Bohn T. Interaction of divalent minerals with liposoluble nutrients and phytochemicals during digestion and influences on their bioavailability – a review. Food Chem. 2018;252:285-93.
17. Calvo MM, Bujanda L, Heras I, Calderon A, Cabriada JL, Orive V, et al. Magnetic resonance cholangiography versus ultrasound in the evaluation of the gallbladder. J Clin Gastroenterol. 2002;34:233-6.
18. Malik AA, Wani ML, Tak SI, Irshad I, Ul-Hassan N. Association of dyslipidaemia with cholilithiasis and effect of cholecystectomy on the same. Int J Surg. 2011;9:641-2.
19. World Health Organization. Obesity and overweight [Internet]. Geneva: World Health Organization; Available from: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight
20. Shi Y, Zou Y, Shen Z, Xiong Y, Zhang W, Liu C, et al. Trace Elements, PPARs, and metabolic syndrome. Int J Mol Sci. 2020;21:2612.
21. Cao J, Mao W, Jiang Q, Huang X, Tan Y, Zuo F, et al. Hepatic steatosis mediates the relationship between cholecystectomy and BMI increase: a population-based study. Front Med (Lausanne). 2025;12:1620036.
22. Huh JH, Lee KJ, Cho YK, Moon S, Kim YJ, Han KD, et al. Cholecystectomy increases the risk of metabolic syndrome in the Korean population: a longitudinal cohort study. Hepatobiliary Surg Nutr. 2023;12:523-33.
23. Chen Y, Wu S, Tian Y. Cholecystectomy as a risk factor of metabolic syndrome: from epidemiologic clues to biochemical mechanisms. Lab Invest. 2018;98:7-14.
24. Bahri MH, Navabian S, Akbari H, Bagherpour JZ, Bozorgmehr R, Mohammaditabar M. Impact of cholecystectomy on metabolic dysfunction-associated steatotic liver disease and metabolic syndrome: a 6-month prospective cohort study. BMC Gastroenterol. 2025;25:671.

25. Tinkov AA, Korobeinikova TV, Morozova GD, Aschner M, Mak DV, Santamaria A, et al. Association between serum trace element, mineral, and amino acid levels with non-alcoholic fatty liver disease (NAFLD) in adult women. J Trace Elem Med Biol. 2024;83:127397.
26. Wang S, Wang Y, Xu J, Chen Y. Is the oral contraceptive or hormone replacement therapy a risk factor for cholelithiasis: a systematic review and meta-analysis. Medicine (Baltimore). 2017;96:e6556.
27. Grundy K, Nelson A. Gallbladder disease. InnovAiT. 2025;18:24-30.
28. Ning Q, Liu F, Fang Y, Zhu X, Liu J, Li Z. Estimating global prevalence of gallbladder stones in general population from 2000 to 2024: systematic review and meta-analysis. Ann Med. 2025;57:2570795.
29. Yan S, Yu L, Fang S, Gu C. The association between intakes of dietary trace minerals and gallstone disease: a cross-sectional study from National Health and Nutrition Examination Survey 2017 to 2018. Medicine (Baltimore). 2024;103:e37741.
30. Nishito Y, Kambe T. Absorption mechanisms of iron, copper, and zinc: an overview. J Nutr Sci Vitaminol (Tokyo). 2018;64:1-7.
31. Einhorn V, Haase H, Maares M. Interaction and competition for intestinal absorption by zinc, iron, copper, and manganese at the intestinal mucus layer. J Trace Elem Med Biol. 2024;84:127459.
32. Stojančević M, Pavlović N, Goločorbin-Kon S, Mikov M. Application of bile acids in drug formulation and delivery. Front Life Sci. 2013;7:112-22.
33. Amaral Raposo M, Sousa Oliveira E, Dos Santos A, Guadagnini D, El Mourabit H, Housset C, et al. Impact of cholecystectomy on the gut-liver axis and metabolic disorders. Clin Res Hepatol Gastroenterol. 2024;48:102370.
34. Ahmed M. Functional, diagnostic and therapeutic aspects of bile. Clin Exp Gastroenterol. 2022;15:105-20.
35. Naik S, Abuji K, Dahiya D, Sharma P, Das R, Behera A, et al. Prevalence of iron deficiency anaemia in patients of cholelithiasis undergoing laparoscopic cholecystectomy. Int Surg J. 2022;9:1335-9.
36. Ciosek Ż, Kot K, Rotter I. Iron, zinc, copper, cadmium, mercury, and bone tissue. Int J Environ Res Public Health. 2023;20:2197.
37. Bárány E, Bergdahl IA, Bratteby LE, Lundh T, Samuelson G, Skerfving S, et al. Iron status influences trace element levels in human blood and serum. Environ Res. 2005;98:215-23.
38. Fujiwara Y, Lee JY, Banno H, Imai S, Tokumoto M, Hasegawa T, et al. Cadmium induces iron deficiency anemia through the suppression of iron transport in the duodenum. Toxicol Lett. Toxicol Lett. 2020;332:130-9.
39. Schuchardt JP, Hahn A. Intestinal absorption and factors influencing bioavailability of magnesium-an update. Curr Nutr Food Sci. 2017;13:260-78.
40. Kiela PR, Ghishan FK. Physiology of intestinal absorption and secretion. Best Pract Res Clin Gastroenterol. 2016;30:145-59.
41. Vijayvargiya P, Camilleri M, Carlson P, Nair A, Nord SL, Ryks M, et al. Effects of colesevelam on bowel symptoms, biomarkers, and colonic mucosal gene expression in patients with bile acid diarrhea in a randomized trial. Clin Gastroenterol Hepatol. 2020;18:2962-70.e6.
42. Porr PJ, Szántay J, Rusu M. Post-cholecystectomy syndrome and magnesium deficiency. J Am Coll Nutr. 2004;23:745S-7S.
43. Fraser WD, Taggart DP, Fell GS, Lyon TD, Wheatley D, Garden OJ, et al. Changes in iron, zinc, and copper concentrations in serum and in their binding to transport proteins after cholecystectomy and cardiac surgery. Clin Chem. 1989;35:2243-7.
44. Wang X, Zhong Y, Zhu Z, Zhang N, Chen X, Wang F, et al. Gut microbiota: a new perspective for bioavailability of selenium and human health. Npj Sci Food. 2025;9:228.
45. Ferreira RLU, Sena-Evangelista KCM, de Azevedo EP, Pinheiro FI, Cobucci RN, Pedrosa LFC. Selenium in human health and gut microflora: bioavailability of selenocompounds and relationship with diseases. Front Nutr. 2021;8:685317.
46. Ha HY, Alfulaij N, Berry MJ, Seale LA. From selenium absorption to selenoprotein degradation. Biol Trace Elem Res. 2019;192:26-37.
47. GarcÍa E, Cabrera C, Lorenzo ML, LÓpez MC, SÁnchez J. Estimation of chromium bioavailability from the diet by an in vitro method. Food Addit Contam. 2001;18:601-6.
48. Sarı İ, Bolat H, Özmen E, Köse B, Ayan D. Effect of laparoscopic cholecystectomy on plasma levels of EETs, arachidonic acids derived lipid mediators. Postgrad Med J. 2025;101:828-35.
49. Ticho AL, Malhotra P, Dudeja PK, Gill RK, Alrefai WA. Intestinal absorption of bile acids in health and disease. Compr Physiol. 2019;10:21-56.
50. Rodríguez-Antonio I, López-Sánchez GN, Garrido-Camacho VY, Uribe M, Chávez-Tapia NC, Nuño-Lámbarri N. Cholecystectomy as a risk factor for non-alcoholic fatty liver disease development. HPB. 2020;22:1513-20.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Case Report
DOI: 10.4274/eajem.galenos.2026.07673
Eurasian J Emerg Med. 2026;25: 489-93

Fever after Percutaneous Cholecystostomy Catheter Removal: Gallstone Extrusion Through a Persistent Cholecystostomy Tract

Shun Yee Low, Corinne Yee Lyn Lau, Chim Yoong Tan, Eunizar Omar

Sengkang General Hospital, Clinic of Emergency Medicine, SingHealth, Singapore

Abstract

Gallstone extrusion through a persistent percutaneous cholecystostomy tract is an unusual complication following catheter removal. We report a 72-year-old institutionalised male who re-attended the emergency department with three days of fever without abdominal pain, tenderness, leukocytosis, or liver biochemical abnormalities. Six days earlier, a partially dislodged percutaneous cholecystostomy catheter, inserted six months prior for acute gangrenous cholecystitis, was removed after unsuccessful exchange while he remained clinically well. Two days before re-attendance, the catheter site was unremarkable. On re-attendance, catheter site inspection revealed extrusion of gallstone-like material. Computed tomography demonstrated acute-on-chronic cholecystitis with an inflamed tract extending from the gallbladder fundus towards the abdominal wall. Gross examination identified gallstones, and fluoroscopic contrast injection demonstrated communication between the cutaneous tract and gallbladder. A new percutaneous cholecystostomy catheter was advanced through the tract for drainage and source control. Given the prolonged catheterisation, short interval after catheter removal, and fluoroscopic findings, persistence of the mature tract was considered more likely than a new cholecystocutaneous fistula, although the latter could not be excluded. This complication may present predominantly with fever and few localising findings. In patients with limited history, reassessment of the previous procedural site may provide an important diagnostic clue and facilitate timely management.

Keywords: Gallstones, cholecystostomy, cholecystitis, fever, aged, emergency service, hospital

Introduction

Percutaneous cholecystostomy is used as a temporising or definitive treatment for acute cholecystitis in high-risk surgical candidates, but complications include catheter blockage, bile leakage, infection, and dislodgement (1). During prolonged catheterisation, the percutaneous tract generally matures within 3-4 weeks of catheter placement (1). Isolated case reports have described tract-related complications following percutaneous drainage for gallbladder disease, including gallstone migration, impaction or extrusion (2-6), and cholecystocutaneous fistula formation (3-8).

We report a 72-year-old institutionalised male who re-attended the emergency department (ED) six days after removal of the percutaneous cholecystostomy catheter, presenting with fever and few localising features. Gallstones were subsequently found extruding through a persistent percutaneous cholecystostomy tract.

Case Report

A 72-year-old male had a history of acute gangrenous cholecystitis that was treated with radiology-guided percutaneous cholecystostomy in October 2025. Two months later, he was admitted for acute cholangitis complicated by *Klebsiella* bacteraemia and underwent endoscopic retrograde cholangiopancreatography with sphincterotomy. He received culture-directed antibiotics and was discharged with the percutaneous cholecystostomy catheter *in situ*. At the routine follow-up two months later, the catheter was exchanged. His other medical history included a basal ganglia haemorrhage and schizophrenia. He was minimally communicative, chair-bound, and required assistance with activities of daily living.

Corresponding Author: Shun Yee Low MD, Sengkang General Hospital, Clinic of Emergency Medicine, SingHealth, Singapore
E-mail: low.shun.yee@singhealth.com.sg **ORCID ID:** orcid.org/0000-0001-6606-4806

Cite this article as: Low SY, Lau CYL, Tan CY, Omar E. Fever after percutaneous cholecystostomy catheter removal: gallstone extrusion through a persistent cholecystostomy tract. Eurasian J Emerg Med. 2026;25: 489-93

Received: 02.07.2026
Accepted: 06.10.2026
Published: 26.10.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Six days before ED re-attendance, the percutaneous cholecystostomy catheter was found to be partially dislodged during routine outpatient follow-up. Catheter exchange was attempted but unsuccessful. As there was minimal drainage and he remained clinically well without fever, abdominal pain, or significant computed tomography (CT) evidence of biliary complications, the Interventional Radiology and General Surgery teams decided to remove the catheter without further biliary intervention. Granulation tissue at the catheter exit site was treated with silver nitrate, and follow-up was arranged.

Four days after catheter removal, he presented to the ED from his nursing home with one day of reported fever, with temperatures up to 37.9 °C, associated with myalgia and lethargy but without abdominal pain. His temperature in the ED was 37.6 °C, and his other vital signs were normal. The examination revealed a soft, non-tender abdomen. The previous percutaneous cholecystostomy site in the right upper quadrant was documented as unremarkable, without erythema, bleeding, or discharge. White blood cell (WBC) count was 9.18 $\times 10^9$/L, total bilirubin was 6 µmol/L, alkaline phosphatase (ALP) was 117 U/L, alanine transaminase (ALT) was 16 U/L, and aspartate transaminase (AST) was 22 U/L. Chest radiography, urinalysis, and SARS-CoV-2 antigen testing did not identify an alternative source of infection. As he remained clinically well, he was discharged with a presumed viral illness and advice for the nursing home staff to monitor him for persistent fever or new symptoms.

Two days later, he re-attended the ED with a three-day history of fever and chills and a maximum temperature of 39 °C. No abdominal pain, jaundice, pale stools, or tea-coloured urine were present. He was afebrile in the ED with normal vital signs. Re-examination revealed a soft, non-tender abdomen. Closer inspection of the prior percutaneous cholecystostomy tract in the right upper quadrant revealed material suspicious for gallstones extruding from the wound, with surrounding erythema (Figure 1). WBC count was 7.80 $\times 10^9$/L, total bilirubin was 3 µmol/L, ALP was 111 U/L, ALT was 17 U/L, and AST was 30U/L. Procalcitonin was 0.11 µg/L (reference range ≤0.49 µg/L), while C-reactive protein (CRP) was elevated at 217 mg/L (reference range ≤4.9 mg/L). Blood cultures were obtained, and intravenous piperacillin-tazobactam (4.5 g every 6 hours) was commenced empirically, based on previous blood culture susceptibilities. Repeat chest radiography and urinalysis were unrevealing. CT of the abdomen and pelvis demonstrated acute-on-chronic cholecystitis with an inflamed tract extending from the gallbladder fundus to the superficial right upper anterior abdominal wall, likely related to the prior percutaneous cholecystostomy tract (Figure 2 A-C). There was no dense gallstone, intraductal calculus, biliary ductal dilatation, or large loculated collection.

The working diagnosis was acute-on-chronic cholecystitis with gallstone extrusion through a persistent percutaneous cholecystostomy tract. On gross examination, the extruding material was identified by the General Surgery team as gallstones. No histopathological confirmation was documented. Given his high operative risk, he underwent radiology-guided reinsertion of a percutaneous cholecystostomy catheter two days after admission. The cutaneous opening at the previous catheter site was cannulated, and contrast was injected under fluoroscopic guidance, demonstrating communication between the tract and the gallbladder (Figure 2D). A new catheter was advanced through the tract into the gallbladder for biliary drainage and source control. Purulent material was aspirated, and bile was collected aseptically for aerobic culture.

Blood cultures remained negative, while bile cultures grew *Staphylococcus aureus* and *Proteus mirabilis*. After two days of intravenous piperacillin-tazobactam, antimicrobial therapy was changed, based on bile culture susceptibilities, to intravenous amoxicillin/clavulanate 1.2 g every 6 hours for three days. The cutaneous opening continued to drain small amounts of haemoserous fluid.



Figure 1. Gallstones extruding from the previous percutaneous cholecystostomy catheter site in the right upper quadrant of the abdomen

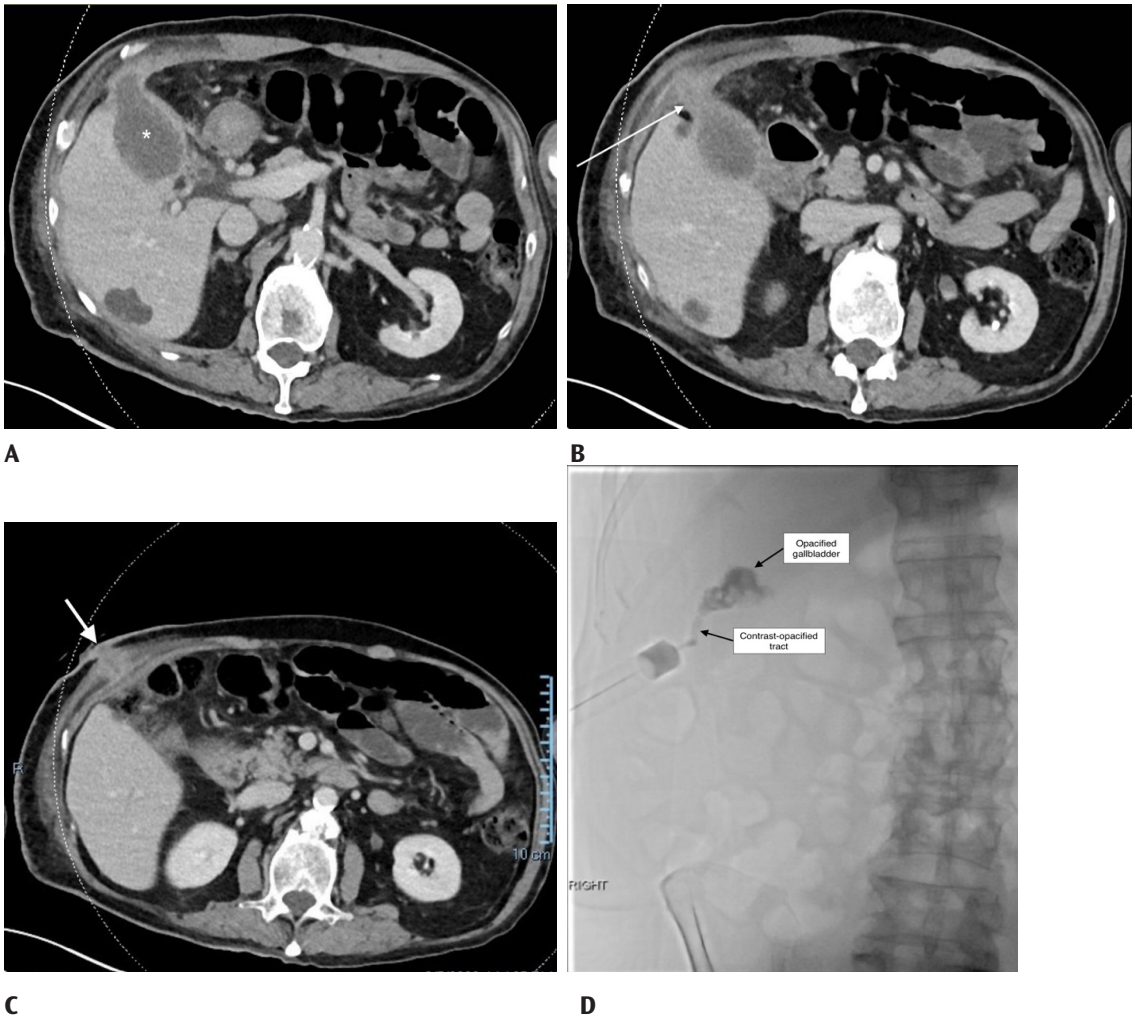


Figure 2. A. An axial contrast-enhanced computed tomography image showing a distended gallbladder (white asterisk) with diffuse mural thickening and pericholecystic fat stranding, consistent with acute-on-chronic cholecystitis. **B.** An axial contrast-enhanced computed tomography image demonstrating the proximal portion of an inflamed tract extending from the gallbladder fundus towards the right upper anterior abdominal wall (thin white arrow). **C.** An axial contrast-enhanced computed tomography image showing the distal portion of the tract extending to the superficial right upper anterior abdominal wall at the previous percutaneous cholecystostomy catheter site (thick white arrow). **D.** A fluoroscopic image obtained during percutaneous cholecystostomy catheter reinsertion following cannulation of the cutaneous opening at the previous catheter site and contrast injection. The labelled contrast-opacified tract communicates with the opacified gallbladder

Three days after the procedure, he had clinically improved and was afebrile, with CRP having decreased to 101 mg/L. He was discharged the following day with the catheter *in situ* and oral amoxicillin/clavulanate 875/125 mg twice daily for four days to complete a seven-day course of amoxicillin/clavulanate. At one- and three-month follow-up visits, he remained well without recurrent fever or cholecystitis. The catheter remained functional with scheduled exchanges every three months. The long-term status of the separate cutaneous opening was not specifically documented.

Discussion

Previous reports of gallstone migration, impaction, or extrusion through prior percutaneous drainage tracts for gallbladder disease, including cases associated with cholecystocutaneous fistulas, have described localising features such as abdominal pain or discomfort (3-5), pain at the previous surgical site (6), and purulent discharge (3-5). In contrast, this patient presented predominantly with fever, without abdominal pain or tenderness. At re-attendance, pain at the previous catheter site

and purulent tract discharge were not specifically documented, and therefore their absence cannot be confirmed. WBC count and liver biochemical tests were normal. The limited documented localising features, together with the absence of leukocytosis or liver biochemical abnormalities, highlight the diagnostic challenge of this complication in the ED.

Recognition was further complicated by the patient's limited ability to provide a reliable history, making symptom-based assessment more challenging. At initial ED attendance, the catheter site was documented as unremarkable, without erythema, bleeding, or discharge. Two days later, however, gallstone extrusion was evident on re-attendance, suggesting interval evolution of the local findings, although an initially subtle abnormality cannot be excluded. This case reinforces the value of repeat examination on ED re-attendance, including inspection of recent procedural sites, when clinically relevant, particularly when history-taking is limited. Clear advice to caregivers to return if persistent fever or changes occur at procedural sites may facilitate earlier reassessment.

The nature of the gallbladder-to-skin tract also warranted consideration. Cholecystocutaneous fistula formation following percutaneous drainage for gallbladder disease has recently been described, including a case in which a drainage catheter for a pericholecystic abscess dislodged after seven days and a fistula became apparent eight months later (8). In contrast, our patient's percutaneous cholecystostomy catheter had remained *in situ* for six months and had been removed only six days before re-attendance. CT demonstrated an inflamed tract extending from the gallbladder fundus towards the abdominal wall; fluoroscopic contrast injection confirmed communication between the cutaneous opening and the gallbladder. These findings could not definitively distinguish persistence of the pre-existing mature cholecystostomy tract from a newly formed cholecystocutaneous fistula. Given the prolonged catheterisation, short interval following catheter removal, and continuity with the previous catheter site, persistence of the mature iatrogenic tract was considered more likely. We therefore describe this finding as gallstone extrusion through a persistent percutaneous cholecystostomy tract, while acknowledging that a newly formed cholecystocutaneous fistula could not be excluded.

An elevated CRP, CT evidence of acute-on-chronic cholecystitis and tract inflammation, a purulent aspirate, and clinical improvement following biliary drainage and antimicrobial therapy supported a probable biliary source for the fever. Normal bilirubin, ALP, and transaminase levels were compatible with gallbladder inflammation in the absence of common bile duct obstruction, consistent with the lack of biliary dilatation or intraductal calculi on CT. Blood cultures remained negative. Although bile cultures grew *Staphylococcus aureus* and *Proteus mirabilis,* these findings should be interpreted cautiously, given that the sample was obtained after the initiation of antimicrobial therapy and prolonged biliary catheterisation may have resulted in biliary colonisation. Although bedside ultrasonography was not performed, it may identify features of acute cholecystitis such as gallstones, gallbladder wall thickening, and pericholecystic fluid (9), potentially aiding earlier recognition in the ED.

Management should be individualised according to clinical status, operative risk, disease severity, and goals of care. Cholecystectomy may be considered in surgically suitable patients, whereas percutaneous drainage may be appropriate for those at high operative risk (1). Given the patient's high operative risk, definitive surgical management was not pursued. Reinsertion of a percutaneous cholecystostomy catheter through the tract provided biliary drainage and source control. He remained well, without recurrent fever or cholecystitis, during three months of follow-up, with planned three-monthly catheter exchanges.

Conclusion

Gallstone extrusion through a persistent percutaneous cholecystostomy tract may present primarily with fever and minimal localising findings. In patients who re-attend after recent catheter removal, particularly when history-taking is limited, repeat examination and inspection of the previous procedural site may reveal an important diagnostic clue. Early recognition may facilitate appropriate investigation and timely source control.

Ethics

Informed Consent: The patient lacked mental capacity to provide informed consent. Written informed consent for publication of this anonymised case report and accompanying clinical images was obtained from the patient's legal guardian.

Footnotes

Author Contributions

Surgical and Medical Practices: C.Y.L.L., C.Y.T., Concept: S.Y.L., E.O., C.Y.L.L., C.Y.T., Design: S.Y.L., E.O., Data Collection or Processing: S.Y.L., E.O., C.Y.L.L., C.Y.T., Analysis or Interpretation: S.Y.L., E.O., C.Y.L.L., C.Y.T., Literature Search: S.Y.L., E.O., Writing: S.Y.L., E.O., C.Y.L.L., C.Y.T.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Hung YL, Sung CM, Fu CY, Liao CH, Wang SY, Hsu JT, et al. Management of patients with acute cholecystitis after percutaneous cholecystostomy: from the acute stage to definitive surgical treatment. Front Surg. 2021;8:616320.
2. Hariharan D, Lobo DN. Spontaneous extrusion of gallstones after percutaneous drainage. Ann R Coll Surg Engl. 2017;99:e1-e2.
3. Qureshi M, Goh B, Strauss P, Nyandowe M. The spontaneous expulsion of a gallstone from a cholecystocutaneous fistula following percutaneous cholecystostomy. Cureus. 2025;17:e79282.
4. Narvaez V, Lin A, Kapil A, Kartiko S. Cholecystocutaneous fistula after percutaneous cholecystostomy tube removal. ACS Case Rev Surg. 2021;3: 93-6.
5. Pripotnev S, Petrakos A. Cholecystocutaneous fistula after percutaneous gallbladder drainage. Case Rep Gastroenterol. 2014;8:119-22.
6. Naim AJ, De Robles MS. Cholecystostoma: unusual evolution of a cholecystocutaneous fistula following percutaneous incision and drainage of perforated cholecystitis. J Surg Case Rep. 2023;2023:rjad477.
7. Lofgren DH, Vasani S, Singzon V. Abdominal wall abscess secondary to cholecystocutaneous fistula via percutaneous cholecystostomy tract. Cureus. 2019;11:e4444.
8. Huggett JJ, Anosike C, Oldfield FE, Sarfraz N. Cholecystocutaneous fistula complicating percutaneous drainage of a pericholecystic abscess, managed conservatively. J Surg Case Rep. 2026;2026:rjag799.
9. Huang SS, Lin KW, Liu KL, Wu YM, Lien WC, Wang HP. Diagnostic performance of ultrasound in acute cholecystitis: a systematic review and meta-analysis. World J Emerg Surg. 2023;18:54.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.64022
Eurasian J Emerg Med. 2026;25: 494-501

Evaluating the Readability, Quality, and Reliability of Responses Generated by ChatGPT, Gemini, and Perplexity to the Most Frequently Searched Queries About Acute Mesenteric Ischemia

Sema Ayten[1], Nihal Dolanbay[2]

[1]Göztepe Prof. Dr. Süleyman Yalçın City Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye
[2]SANKO University Faculty of Health Sciences, Department of Surgical Nursing, Gaziantep, Türkiye

Abstract

Aim: Patients increasingly consult artificial intelligence (AI) chatbots before deciding whether to seek emergency care. This study compared the readability, reliability, and quality of responses generated by ChatGPT, Gemini, and Perplexity in response to frequently searched queries about acute mesenteric ischemia.

Materials and Methods: Sixteen queries identified through Google Trends were submitted to each chatbot, yielding 48 responses. Readability was assessed with seven formulas. Two evaluators independently rated reliability using the JAMA benchmarks and modified DISCERN and quality using the Global Quality score and the Ensuring Quality Information for Patients tool. Platforms were compared with the Friedman and Wilcoxon signed-rank tests.

Results: ChatGPT produced shorter responses than Gemini and Perplexity (medians 180.5, 390.0, and 416.5 words; $p<0.001$). All platforms' content was written at a college reading level (median Flesch-Kincaid Grade Level of 14.0-15.5), well above the recommended sixth to eighth-grade level. Perplexity obtained the highest JAMA and Ensuring Quality Information for Patients scores, whereas Gemini and Perplexity outperformed ChatGPT on the modified DISCERN and the Global Quality score.

Conclusion: No chatbot was superior across all measures, and none produced text readable by the average patient. Information on this time-critical emergency should be simplified and professionally reviewed before being recommended to patients. The clinical accuracy and the appropriateness of emergency referral advice were not assessed and require further evaluation.

Keywords: Mesenteric ischemia, artificial intelligence, chatbots, health literacy, readability

Introduction

Acute mesenteric ischemia (AMI; not to be confused with acute myocardial infarction) is an uncommon but highly lethal abdominal emergency in which an abrupt reduction in intestinal blood flow leads, if untreated, to bowel necrosis, multiorgan failure, and death (1-3). The condition comprises arterial embolism, arterial thrombosis, mesenteric venous thrombosis, and non-occlusive mesenteric ischemia, because each follows a different clinical course, the presentation is often nonspecific and the diagnosis is frequently delayed (3-5). No laboratory biomarker is accurate enough to establish the diagnosis on its own (6), and computed tomography angiography remains the cornerstone of evaluation (3,7). Management ranges from endovascular or open revascularization to bowel resection, depending on the mechanism and bowel viability (1,8,9), and every delay increases bowel loss and mortality (3).

Artificial intelligence (AI) has so far been explored in AMI mainly as a clinician-facing tool, for example to predict mortality or to interpret computed tomography angiography images (4,10,11).

Corresponding Author: Sema Ayten MD, Göztepe Prof. Dr. Süleyman Yalçın City Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye
E-mail: semayten@hotmail.com **ORCID ID:** orcid.org/0000-0001-8417-331X

Cite this article as: Ayten S, Dolanbay N. Evaluating the readability, quality, and reliability of responses generated by ChatGPT, Gemini, and Perplexity to the most frequently searched queries about acute mesenteric ischemia. Eurasian J Emerg Med. 2026;25: 494-501

Received: 28.09.2026
Accepted: 15.10.2026
Published: 27.10.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

A different and far less-studied application is the patient-facing chatbot. Large language model-based platforms such as ChatGPT, Gemini, and Perplexity now answer health questions in conversational language and are increasingly evaluated as sources of medical information (12-14). Their answers, however, vary considerably in reliability and readability, and some appear convincing while conflicting with clinical practice guidelines (12-15).

Because patients may consult these platforms before seeking professional care (16), the information they provide must be understandable, reliable, and clinically appropriate. Studies across several specialties have shown that chatbot responses frequently exceed the reading level recommended for patient materials (14,17-19). Structured instruments allow these characteristics to be assessed systematically: the JAMA benchmarks examine authorship, attribution, disclosure, and currency (20); DISCERN evaluates the reliability of consumer health information (21); the Global Quality score (GQS) rates overall quality and usefulness (22); and the Ensuring Quality Information for Patients (EQIP) tool assesses the content, identification, and structure of written patient information (23).

For a time-critical emergency such as AMI, the practical question is not only whether chatbot information is readable and reliable, but also whether it helps patients recognize that urgent evaluation is needed. To our knowledge, chatbot-generated information on AMI has not been evaluated previously.

This study aimed to compare the readability, reliability, and quality of responses generated by ChatGPT, Gemini, and Perplexity in response to the most frequently searched queries about AMI.

Materials and Methods

Study Design

This comparative cross-sectional study evaluated the information generated by three AI platforms. The evaluated platforms were ChatGPT, Gemini, and Perplexity. The study focused on patient-oriented information about AMI. No patients, medical records, or personally identifiable information were included. The study was reported in accordance with the Chatbot Assessment Reporting Tool guideline, where applicable (24).

Selection of Search Queries

Frequent search queries related to AMI were identified using Google Trends (Google LLC, Mountain View, CA, USA) on 10 July 2026. The search was conducted using "AMI" as the initial search term, covering the preceding five years, with the geographical setting worldwide, without country-specific restrictions. Related queries were examined under the "Top" category, and the 16 highest-ranked eligible queries were selected for analysis. Duplicate or overlapping queries and those without a plausible clinical relationship to AMI were excluded. Queries were retained in their original keyword format rather than converted into questions to reflect actual public search behavior, consistent with previous studies (12,14).

Broader terms, including "Thrombosis," "Acute Abdomen," and "Peritonitis," were retained because they appeared among the top-ranked related queries and represented clinically relevant concepts associated with the vascular mechanisms, clinical presentation, or complications of AMI. Although these terms are not specific to AMI, their inclusion was intended to reflect broader information-seeking behavior related to the condition. All three platforms interpreted the abbreviated query "SMA" in the context of the superior mesenteric artery.

1. Acute mesenteric ischemia
2. Mesenteric ischemia treatment
3. Mesenteric ischemia symptoms
4. Mesenteric artery
5. Acute mesenteric ischemia symptoms
6. Thrombosis
7. Superior mesenteric artery
8. SMA
9. Acute abdomen
10. Intestinal ischemia
11. Peritonitis
12. Mesenteric ischemia radiology
13. Mesenteric venous thrombosis
14. Bowel infarction
15. Bowel necrosis
16. Mesenteric ischemia diagnosis

Artificial Intelligence Response Collection

Responses were collected on 10 July 2026 through the official web interfaces of ChatGPT (OpenAI, San Francisco, CA, USA), Gemini (Google LLC, Mountain View, CA, USA), and Perplexity (Perplexity AI, San Francisco, CA, USA). All three platforms were accessed using their freely available versions without requiring user login.

These platforms use proprietary, closed-source models. Default settings were maintained, and no specific model was manually selected. Because the interfaces did not provide independently verifiable model identifiers for each response, the exact underlying model versions could not be confirmed. Perplexity was accessed using its default automatic model-selection mode.

Each of the 16 queries was entered in English, exactly as retrieved from Google Trends, in a separate, new chat session without prior context, additional instructions, or prompt modification. Only the first response generated for each query was recorded, yielding 48 responses for analysis. No repeated queries or follow-up prompts were used.

The sample size was determined by the 16 highest-ranked eligible queries identified through Google Trends, rather than by a formal power calculation. Each query was submitted once to each of the three platforms, resulting in 48 responses. This approach was intended to provide a standardized cross-sectional comparison of the platforms under identical query conditions at a single point in time, rather than to assess within-platform response variability.

Readability Assessment

Readability was calculated using the Readability Formulas online tool (https://readabilityformulas.com). Seven measures were computed for each response: the Flesch Reading Ease score (FRES) (25), Flesch-Kincaid Grade Level (FKGL) (26), Gunning Fog index (27), Simple Measure of Gobbledygook (SMOG) (28), Automated Readability index (ARI) (29), Coleman-Liau index (30), and Linsear Write Formula. FRES ranges from 0 to 100, with higher scores indicating easier text; the remaining measures estimate the United States school grade level required to understand the text. A sixth- to eighth-grade level is generally recommended for patient education materials (31). Before readability analysis, all chatbot responses underwent standardized text preprocessing. URLs, citations, reference markers, headings, bullet-point symbols, and other non-prose elements were removed. The remaining textual content was analyzed using the seven readability formulas. The same preprocessing procedure was applied consistently across all three platforms to minimize potential formatting-related differences in readability scores.

Quality and Reliability Assessment

Reliability was assessed with the JAMA benchmarks (20) and the modified DISCERN (21,32). The JAMA benchmarks award one point each for authorship, attribution, disclosure, and currency (total score 0-4). The modified DISCERN consists of five yes/no questions addressing clarity, reliability of sources, balance, provision of additional sources, and reference to areas of uncertainty (total score 0-5). Overall quality was rated using the GQS, a five-point scale ranging from 1 (poor) to 5 (excellent) (22). Patient information quality was assessed with the EQIP tool (23). The original 20-item version (EQIP-20) was used: each item was rated as “yes”, “partly”, “no”, or “not applicable”, and the total score was expressed as a percentage, with higher scores indicating better quality.

All responses were independently rated using the four instruments by two evaluators with different professional backgrounds: an emergency medicine specialist and a faculty member holding a PhD in surgical nursing. Because Perplexity displays source links within its responses, complete blinding to the platform was not feasible. Inter-rater agreement was calculated from the independent ratings; disagreements were then resolved by discussion, and the consensus scores were used in all comparative analyses.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, version 27.0 (IBM Corp., Armonk, NY, USA). Continuous variables are presented as median (minimum–maximum) and 25^{th}-75^{th} percentiles. Differences among the three platforms were assessed using the Friedman test, and pairwise comparisons were evaluated using the Wilcoxon signed-rank test. A Bonferroni correction was applied to the three pairwise comparisons; therefore, $p<0.0167$ was considered significant for these tests, while $p<0.05$ was considered significant for other tests. Effect sizes for the Friedman test were expressed as Kendall’s W, calculated as $\chi^2/[N(k-1)]$, where N is the number of queries and k the number of platforms; values of approximately 0.1, 0.3, and 0.5 were interpreted as small, moderate, and large effects, respectively.

Inter-rater reliability was evaluated using intraclass correlation coefficients (ICCs) based on a two-way mixed-effects model with absolute agreement; single-measure ICC values with 95% confidence intervals (CIs) are reported. ICC values below 0.50, between 0.50 and 0.75, between 0.75 and 0.90, and above 0.90 were interpreted as poor, moderate, good, and excellent agreement, respectively (33).

Results

A total of 48 responses from three AI platforms were evaluated. Each platform generated responses to the same 16 AMI-related queries. Effect sizes for the overall comparisons ranged from moderate (Kendall’s W=0.31 for Linsear Write) to large (Kendall’s W=1.00 for JAMA), and are presented in Table 1.

Table 1. Overall comparison of readability, reliability, and quality measures among ChatGPT, Gemini, and Perplexity

Parameter	ChatGPT mean rank	Gemini mean rank	Perplexity mean rank	χ^2	df	p	Kendall's W
Word count	1.13	2.38	2.50	18.500	2	<0.001	0.58
FRES	1.06	2.78	2.16	25.410	2	<0.001	0.79
FKGL	2.16	1.31	2.53	12.667	2	0.002	0.40
Gunning Fog index	2.75	1.19	2.06	19.625	2	<0.001	0.61
SMOG	1.13	2.00	2.88	24.500	2	<0.001	0.77
ARI	1.44	1.56	3.00	24.125	2	<0.001	0.75
Coleman-Liau index	2.00	1.38	2.63	12.500	2	0.002	0.39
Linsear Write Formula	2.63	1.81	1.56	9.875	2	0.007	0.31
JAMA	1.50	1.50	3.00	32.000	2	<0.001	1.00
Modified DISCERN	1.22	2.47	2.31	19.000	2	<0.001	0.59
GQS	1.50	2.16	2.34	12.182	2	0.002	0.38
EQIP	1.28	2.16	2.56	15.404	2	<0.001	0.48

Friedman test. χ^2: chi-square, df: degrees of freedom, FRES: Flesch Reading Ease Score, FKGL: Flesch-Kincaid Grade Level, SMOG: Simple Measure of Gobbledygook, ARI: Automated Readability index, JAMA: Journal of the American Medical Association benchmarks, GQS: Global Quality score, EQIP: Ensuring Quality Information for Patients. Kendall's $W=\chi^2/[N(k-1)]$, N=16 queries, k=3 platforms

ChatGPT produced a median word count of 180.50 words. Median word counts were 390.00 for Gemini and 416.50 for Perplexity. Overall word counts differed significantly among the platforms (p<0.001). ChatGPT responses were shorter than Gemini responses (p=0.001). They were also shorter than Perplexity responses (p=0.001). Gemini and Perplexity showed no significant difference (p=0.877).

Median FRES scores were 0.00 for ChatGPT, 16.50 for Gemini, and 9.00 for Perplexity. FRES differed significantly among the three platforms (p<0.001). ChatGPT had lower FRES scores than Gemini (p<0.001). ChatGPT also had lower scores than Perplexity (p=0.001). Gemini had higher FRES scores than Perplexity(p=0.014).

FKGL scores differed significantly among the platforms (p=0.002). ChatGPT had higher FKGL scores than Gemini (p=0.013). ChatGPT and Perplexity showed no significant difference (p=0.074). Gemini had lower FKGL scores than Perplexity (p=0.010). The median FKGL values for the three platforms (15.52, 14.04, and 15.47 for ChatGPT, Gemini, and Perplexity) corresponded to a college reading level, exceeding the recommended sixth- to eighth-grade level.

Gunning Fog scores differed significantly among the platforms (p<0.001). ChatGPT scored higher than Gemini (p=0.001). ChatGPT also had higher scores than Perplexity (p=0.006). Gemini had lower scores than Perplexity (p=0.010).

SMOG scores differed significantly among the three platforms (p<0.001). ChatGPT had lower SMOG scores than Gemini (p=0.001). ChatGPT also scored lower than Perplexity (p<0.001). Gemini had lower scores than Perplexity (p<0.001).

ARI scores also differed significantly among the platforms (p<0.001). Perplexity's scores were higher than those of ChatGPT and Gemini (both p<0.001). ChatGPT and Gemini showed no significant difference (p=0.698).

Coleman-Liau scores differed significantly among the platforms (p=0.002). Gemini had significantly lower scores than Perplexity (p=0.004). Other pairwise comparisons remained nonsignificant after Bonferroni correction. Linsear Write scores also differed significantly among platforms (p=0.007). ChatGPT had higher scores than Perplexity (p=0.003). Other pairwise comparisons were not significant after correction. Detailed readability findings are presented in Tables 1 and 2.

JAMA scores differed significantly among the three platforms (p<0.001). ChatGPT and Gemini both had median JAMA scores of 0.00. Perplexity had a median JAMA score of 2.00. Perplexity outperformed both competing platforms (p<0.001 for both comparisons). ChatGPT and Gemini showed no difference (p=1.000).

Modified DISCERN scores also differed significantly among platforms (p<0.001). Median scores were 2.50 for ChatGPT, 4.00 for Gemini, and 3.00 for Perplexity. Gemini scored higher than ChatGPT (p=0.001). Perplexity also scored higher than

ChatGPT (p=0.001). Gemini and Perplexity showed no significant difference (p=0.527).

GQS scores differed significantly among the three platforms(p=0.002). Median scores were 3.00 for ChatGPT and 4.00 for both Gemini and Perplexity. Gemini scored higher than ChatGPT (p=0.008). Perplexity also scored higher than ChatGPT (p=0.003). Gemini and Perplexity showed no significant difference (p=0.414).

EQIP scores differed significantly among the three platforms (p<0.001). Median scores were 47.00 for ChatGPT, 49.00 for Gemini, and 52.00 for Perplexity. Gemini scored higher than GPT (p=0.002). Perplexity also scored higher than ChatGPT (p=0.001). Perplexity scored higher than Gemini (p=0.013). These findings are presented in Tables 1 and 3.

Inter-rater agreement varied across the four instruments. The two raters assigned identical JAMA scores to all 48 responses (100% agreement; ICC=1.000, 95% CI: 1.000-1.000), and EQIP showed good agreement (ICC=0.900, 95% CI: 0.857-0.931). For modified DISCERN (ICC=0.679, 95% CI: 0.272–0.880) and GQS (ICC=0.638, 95% CI: 0.170-0.874), the point estimates indicated moderate agreement, although the wide CIs ranged from poor to good (Table 4).

Discussion

In this study, the three chatbots differed in almost every measure, but none performed best across all of them. Perplexity scored highest on the JAMA benchmarks and EQIP; Gemini and Perplexity scored higher than ChatGPT on modified DISCERN

Table 2. Comparison of word count and readability scores among ChatGPT, Gemini, and Perplexity

Parameter	ChatGPT median (min-max) (25^{th}-75^{th})	Gemini median (min-max) (25^{th}-75^{th})	Perplexity median (min-max) (25^{th}-75^{th})	ChatGPT vs. Gemini, p	ChatGPT vs. Perplexity, p	Gemini vs. Perplexity, p
Word count	180.50 (113-496) [138.50-206.75]	390.00 (211-617) [322.25-458.25]	416.50 (299-576) [324.25-469.00]	0.001	0.001	0.877
FRES	0.00 (0-20) [0.00-5.25]	16.50 (4-56) [9.00-32.00]	9.00 (0-35) [2.75-18.25]	<0.001	0.001	0.014
FKGL	15.52 (12-16) [13.91-15.82]	14.04 (8-21) [11.56-15.16]	15.47 (13-18) [14.87-16.19]	0.013	0.074	0.010
Gunning Fog index	18.25 (16-20) [16.93-19.18]	16.65 (10-18) [14.28-17.38]	17.45 (15-19) [16.50-18.20]	0.001	0.006	0.010
SMOG	8.08 (7-11) [7.74-8.61]	10.81 (8-13) [10.34-11.78]	12.92 (12-14) [12.19-13.44]	0.001	<0.001	<0.001
ARI	13.35 (9-15) [12.60-14.17]	13.35 (9-16) [11.26-14.67]	16.43 (12-19) [14.72-16.82]	0.698	<0.001	<0.001
Coleman-Liau index	17.35 (12-19) [16.36-18.36]	16.07 (11-19) [14.02-17.55]	17.81 (12-21) [15.64-19.07]	0.063	0.171	0.004
Linsear Write Formula	18.34 (11-39) [15.60-28.94]	14.16 (9-23) [12.81-17.12]	13.32 (10-23) [12.14-14.79]	0.017	0.003	0.326

Data are presented as median (minimum-maximum) and [25^{th}-75^{th} percentiles]. Pairwise comparisons: Wilcoxon signed-rank test with Bonferroni correction (p<0.0167 considered significant). FRES: Flesch Reading Ease Score, FKGL: Flesch-Kincaid Grade Level, SMOG: Simple Measure of Gobbledygook, ARI: Automated Readability index, min-max: minimum-maximum

Table 3. Comparison of JAMA, modified DISCERN, GQS, and EQIP scores

Parameter	ChatGPT median (min–max) [25^{th}-75^{th}]	Gemini median (min–max) [25^{th}-75^{th}]	Perplexity median (min–max) [25^{th}-75^{th}]	ChatGPT vs. Gemini, p	ChatGPT vs. Perplexity, p	Gemini vs. Perplexity, p
JAMA	0.00 (0-0) [0.00-0.00]	0.00 (0-0) [0.00-0.00]	2.00 (2-2) [2.00-2.00]	1.000	<0.001	<0.001
Modified DISCERN	2.50 (2-3) [2.00-3.00]	4.00 (3-4) [3.00-4.00]	3.00 (3-4) [3.00-4.00]	0.001	0.001	0.527
GQS	3.00 (3-4) [3.00-4.00]	4.00 (3-4) [3.25-4.00]	4.00 (3-4) [4.00-4.00]	0.008	0.003	0.414
EQIP	47.00 (45-48) [46.00-47.75]	49.00 (45-55) [48.00-50.00]	52.00 (45-56) [49.25-54.00]	0.002	0.001	0.013

Data are presented as median (minimum-maximum) [25^{th}-75^{th} percentiles]. Pairwise comparisons: Wilcoxon signed-rank test with Bonferroni correction (p<0.0167 considered significant). JAMA: Journal of the American Medical Association benchmarks, DISCERN: Modified DISCERN, GQS: Global Quality score, EQIP: Ensuring Quality Information for Patients: min-max: minimum-maximum

Table 4. Inter-rater reliability of the quality and reliability assessment tools

Scoring system	ICC	95% CI	Interpretation
JAMA	1.000	1.000-1.000	Excellent
Modified DISCERN	0.679	0.272-0.880	Moderate
GQS	0.638	0.170-0.874	Moderate
EQIP	0.900	0.857-0.931	Good

Two-way mixed-effects model, absolute agreement, single measures. JAMA scores did not vary within platforms; therefore, the ICC of 1.000 reflects identical ratings on explicit criteria. The wide confidence intervals for modified DISCERN and GQS extend from poor to good agreement. ICC: Interclass correlation coefficient, CI: Confidence interval, GQS: Global Quality score, EQIP: Ensuring Quality Information for Patients

and GQS; ChatGPT produced the shortest responses. All three platforms generated text at a college reading level, which is well above the level recommended for patient information.

Patient education materials are recommended to be written at a sixth- to eighth-grade reading level (31). In our study, median FKGL values ranged from 14.04 to 15.52, and median FRES values between 0 and 16.5 fall within the range classified as very difficult to read (25). This finding is consistent with previous chatbot studies on cardiopulmonary resuscitation, sexually transmitted diseases, and prostate cancer radiotherapy, in which responses also exceeded recommended levels (14,17,18). For AMI, the gap is particularly relevant: a patient with sudden, severe abdominal pain who receives complex, terminology-heavy text may fail to extract the single most important message: that emergency evaluation is needed.

Pairwise readability comparisons depended on the formula used. ChatGPT had the lowest FRES, indicating the most difficult text, yet it also had the lowest SMOG score. This apparent contradiction is more likely to reflect properties of the formulas than a true difference in readability. SMOG was developed for samples of 30 sentences (28), whereas ChatGPT responses were considerably shorter (a median of 180.5 words). Although URLs, citations, reference markers, and other non-prose elements were removed before analysis, differences in vocabulary, sentence structure, and response length may still have contributed to variations among readability formulas. Individual formula results should therefore be interpreted cautiously; the consistent message across formulas is that no platform approached the recommended level.

ChatGPT generated substantially shorter responses than the other two platforms, which may reflect different default response strategies (12-14). Longer answers allow broader explanations, but length alone does not improve understanding and may increase the cognitive burden for a patient with acute symptoms. Perplexity was the only platform to obtain JAMA points, reflecting its practice of displaying source links alongside responses, which corresponds to the attribution and currency criteria of the JAMA benchmarks (20). Similar advantages of Perplexity have been reported in previous comparisons (13,14). However, the presence of citations does not guarantee that the cited sources are accurate, relevant, or even existent. None of the platforms provided authorship or disclosure information.

Gemini and Perplexity outperformed ChatGPT on modified DISCERN, GQS, and EQIP; however, they did not differ from each other on modified DISCERN and GQS. Similarly, a previous study evaluating cardiopulmonary resuscitation training videos on YouTube demonstrated considerable variability in the quality and reliability of digital health information, highlighting the importance of standardized assessment tools and professional oversight (34). These instruments reward balance, discussion of uncertainty, structure, and completeness (21-23), features more likely to be present in longer responses. This may explain why the two platforms with longer responses scored higher, but it also illustrates a trade-off: more complete responses were not easier to read. Comparable platform-related differences have been reported for palliative care, cardiopulmonary resuscitation, microtia, and breast cancer surgery (13,14,19,35).

These findings have specific implications for emergency medicine. Mortality in AMI increases with delays in diagnosis and revascularization (2,5), and structured management pathways aim to shorten this interval (1,8). A patient-facing chatbot is useful in this setting only if it clearly conveys that sudden, severe abdominal pain, particularly in older patients or in those with atrial fibrillation or vascular disease, requires immediate emergency evaluation (1,3). The instruments used in this study do not specifically assess the presence of such urgency messages, and future studies should evaluate this outcome directly. Importantly, higher reliability or quality scores should not be interpreted as evidence of clinical accuracy or guideline adherence. In the context of AMI, the absence of a direct assessment of emergency referral advice and of guideline-consistent recommendations limits the clinical interpretation of these findings.

The AI tools studied in AMI to date, such as mortality prediction models and computed tomography angiography analysis (4,10,11), are designed to support clinicians. Patient-facing chatbots differ fundamentally because their output reaches users without professional filtering. Evidence that chatbot recommendations may conflict with clinical guidelines (15) and that their quality varies across platforms (35) underlines the need for professional oversight.

Although the ICC point estimates indicated moderate inter-rater agreement for modified DISCERN and GQS, the wide 95% CIs suggest considerable uncertainty regarding the precision of these estimates. Therefore, differences between platforms based on these two instruments should be interpreted cautiously, as variability between evaluators may have influenced the observed results. The different professional backgrounds of the two evaluators, emergency medicine and surgical nursing, may also have contributed to the lower agreement for these more subjective instruments, although this combination reflects both the clinical and patient-education perspectives.

Study Limitations

This study has several limitations. First, only three platforms were evaluated; their responses may change with model updates. Second, only one response per query was evaluated. Given the stochastic nature of generative AI, repeated queries may yield different responses. This approach limits the assessment of within-platform variability, reproducibility, and generalizability. Future studies should incorporate repeated queries across multiple time points to evaluate response consistency and stability more comprehensively. Third, queries were entered in the keyword format retrieved from Google Trends. Although this reflects search-engine behavior and is consistent with previous studies (12,14), it may not represent conversational, question-based interactions, which could produce different answers. Some queries were broader terms related to AMI rather than specific to it. Although these broader queries were retained to reflect real-world search behavior, their inclusion may have influenced the observed differences in chatbot performance. Future studies focusing exclusively on AMI-specific queries could help determine whether the present findings remain consistent when broader or ambiguous search terms are excluded. Fourth, only English-language queries were analyzed. Fifth, readability formulas measure linguistic complexity but not actual patient comprehension. Sixth, the quality instruments include subjective components; the two evaluators were from different professional backgrounds; complete blinding of the evaluators to the platform was not feasible because Perplexity displays its sources. Seventh, because the platforms were accessed through free, non-logged-in web interfaces, the exact underlying model versions could not be independently verified; the findings should therefore be interpreted as comparisons of the publicly accessible chatbot platforms under the specified access conditions rather than as comparisons of definitively identified individual language models. Clinical accuracy and guideline adherence were not directly evaluated, including whether responses appropriately recommended an emergency assessment for suspected AMI. Therefore, higher quality or reliability scores should not be interpreted as evidence of clinically accurate or safe recommendations. Future studies should include larger, question-based query sets, repeated sampling, and direct, guideline-based assessment of clinical accuracy and emergency referral advice.

Conclusion

ChatGPT, Gemini, and Perplexity differed in the readability, reliability, and quality of their responses to frequently searched AMI queries, and no platform was superior across all measures. Perplexity provided the most transparent sourcing and the highest quality of patient information, whereas Gemini and Perplexity outperformed ChatGPT in reliability and overall quality. All three platforms, however, produced text well above the recommended reading level. Given the time-critical nature of AMI, these findings highlight the need for more accessible patient information. However, the clinical accuracy, guideline adherence, and appropriateness of emergency referral advice provided by these platforms require further evaluation before their use as reliable sources of patient guidance can be recommended.

Ethics

Ethics Committee Approval: This study evaluated publicly accessible responses generated by artificial intelligence platforms. No human participants or identifiable personal data were included. Therefore, institutional ethics approval was not required.

Informed Consent: This comparative cross-sectional study evaluated the information generated by three AI platforms. The evaluated platforms were ChatGPT, Gemini, and Perplexity. The study focused on patient-oriented information about AMI. No patients, medical records, or personally identifiable information were included.

Footnotes

Authorship Contributions

Concept: S.A., N.D., Design: S.A., N.D., Data Collection or Processing: S.A., N.D., Analysis or Interpretation: S.A., N.D., Literature Search: S.A., N.D., Writing: S.A., N.D.

Conflict of Interest: The authors declare that they have no conflicts of interest.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Bala M, Catena F, Kashuk J, De Simone B, Gomes CA, Weber D, et al. Acute mesenteric ischemia: updated guidelines of the World Society of Emergency Surgery. World J Emerg Surg. 2022;17:54.

2. Tamme K, Reintam Blaser A, Laisaar KT, Mändul M, Kals J, Forbes A, et al. Incidence and outcomes of acute mesenteric ischaemia: a systematic review and meta-analysis. BMJ Open. 2022;12:e062846.
3. Garzelli L, Ben Abdallah I, Nuzzo A, Zappa M, Corcos O, Dioguardi Burgio M, et al. Insights into acute mesenteric ischaemia: an up-to-date, evidence-based review from a mesenteric stroke centre unit. Br J Radiol. 2023;96:20230232.
4. Bourcier S, Klug J, Nguyen LS. Non-occlusive mesenteric ischemia: diagnostic challenges and perspectives in the era of artificial intelligence. World J Gastroenterol. 2021;27:4088-103.
5. Martini V, Lederer AK, Fink J, Chikhladze S, Utzolino S, Fichtner-Feigl S, et al. Clinical characteristics and outcome of patients with acute mesenteric ischemia: a retrospective cohort analysis. Langenbecks Arch Surg. 2022;407:1225-32.
6. Reintam Blaser A, Starkopf J, Björck M, Forbes A, Kase K, Kiisk E, et al. Diagnostic accuracy of biomarkers to detect acute mesenteric ischaemia in adult patients: a systematic review and meta-analysis. World J Emerg Surg. 2023;18:44.
7. Lee MH, Pickhardt PJ, Sorensen AM, Toia GV, Lubner MG, Sappenfield R, et al. Acute mesenteric ischemia: pathophysiology-based approach to imaging findings and diagnosis. Radiographics. 2025;45:e250012.
8. Tilsed JV, Casamassima A, Kurihara H, Mariani D, Martinez I, Pereira J, et al. ESTES guidelines: acute mesenteric ischaemia. Eur J Trauma Emerg Surg. 2016;42:253-70.
9. Shi Y, Zhao B, Zhou Y, Chen L, Su H, Gu J, et al. Endovascular revascularization vs open surgical revascularization as the first strategy for arterial acute mesenteric ischemia: a systematic review and meta-analysis. J Vasc Surg. 2024;80:1883-93.e2.
10. Harmantepe AT, Dulger UC, Gonullu E, Dikicier E, Şentürk A, Eröz E. A method for predicting mortality in acute mesenteric ischemia: machine learning. Ulus Travma Acil Cerrahi Derg. 2024;30:487-92.
11. Song L, Zhang X, Zhang J, Wu J, Wang J, Wang F. Deep learning-assisted diagnosis of acute mesenteric ischemia based on CT angiography images. Front Med (Lausanne). 2025;12:1510357.
12. Gül Ş, Erdemir İ, Hancı V, Aydoğmuş E, Erkoç YS. How artificial intelligence can provide information about subdural hematoma: assessment of readability, reliability, and quality of ChatGPT, BARD, and Perplexity responses. Medicine (Baltimore). 2024;103:e38009.
13. Hancı V, Ergün B, Gül Ş, Uzun Ö, Erdemir İ, Hancı FB. Assessment of readability, reliability, and quality of ChatGPT, BARD, Gemini, Copilot, Perplexity responses on palliative care. Medicine (Baltimore). 2024;103:e39305.
14. Ömür Arça D, Erdemir İ, Kara F, Shermatov N, Odacioğlu M, İbişoğlu E, et al. Assessing the readability, reliability, and quality of artificial intelligence chatbot responses to the 100 most searched queries about cardiopulmonary resuscitation: an observational study. Medicine (Baltimore). 2024;103:e38352.
15. Gianola S, Bargeri S, Castellini G, Cook C, Palese A, Pillastrini P, et al. Performance of ChatGPT compared to clinical practice guidelines in making informed decisions for lumbosacral radicular pain: a cross-sectional study. J Orthop Sports Phys Ther. 2024;54:222-8.
16. Nian PP, Saleet J, Magruder M, Wellington IJ, Choueka J, Houten JK, et al. ChatGPT as a source of patient information for lumbar spinal fusion and laminectomy: a comparative analysis against Google Web Search. Clin Spine Surg. 2024;37:E394-403.
17. Yıldız HA, Söğütdelen E. AI chatbots as sources of STD information: a study on reliability and readability. J Med Syst. 2025;49:43.
18. Ozturk GA, Gurlek Tumer E, Kurt Çoban FS, Ozturk H, Gurlek U, Ozturk M, et al. AI chatbots for patient education in localized prostate cancer radiotherapy: a comparative quality and readability analysis. Patient Educ Couns. 2026;149:109594.
19. Dadi S, Kring T, Latz K, Cohen D, Thaller S. Evaluating the reliability and readability of AI chatbot responses for microtia patient education. J Craniofac Surg. 2026;37:794-7.
20. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet. JAMA. 1997;277:1244-5.
21. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health. 1999;53:105-11.
22. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol. 2007;102:2070-7.
23. Moult B, Franck LS, Brady H. Ensuring quality information for patients: development and preliminary validation of a new instrument. Health Expect. 2004;7:165-75.
24. CHART Collaborative; Huo B, Collins G, Chartash D, Thirunavukarasu A, Flanagin A, Iorio A, et al. Reporting guideline for chatbot health advice studies: the CHART statement. JAMA Netw Open. 2025;8:e2530220.
25. Flesch R. A new readability yardstick. J Appl Psychol. 1948;32:221-33.
26. Kincaid JP, Fishburne RP, Rogers RL, Chissom BS. Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel. Research Branch Report 8-75. Millington (TN): Naval Technical Training Command; 1975.
27. Gunning R. The technique of clear writing. New York: McGraw-Hill; 1952.
28. McLaughlin GH. SMOG grading: a new readability formula. J Reading. 1969;12:639-46.
29. Senter RJ, Smith EA. Automated readability index. AMRL-TR-6620. Wright-Patterson Air Force Base (OH): Aerospace Medical Research Laboratories; 1967.
30. Coleman M, Liau TL. A computer readability formula designed for machine scoring. J Appl Psychol. 1975;60:283-4.
31. Weiss BD. Health literacy and patient safety: help patients understand. Manual for clinicians. 2nd ed. Chicago: American Medical Association Foundation; 2007.
32. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis - a wakeup call? J Rheumatol. 2012;39:899-903.
33. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155-63.
34. Nogay S, Dolanbay T, Demir B, Kartal B, Özkan A. Evaluation of the educational quality and reliability of cardiopulmonary resuscitation training videos on YouTube. Br J Hosp Med (Lond). 2026;87:51256.
35. Naufal M, Kannan S, Senthilkumar S, Palaniappan T, Kumar SS, Kulkarni PP. Patient education using AI: a cross-sectional study comparing ChatGPT and Google Gemini-generated patient education brochures on various surgical management of breast cancer. Cureus. 2025;17:e97525.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Review
DOI: 10.4274/eajem.galenos.2026.38243
Eurasian J Emerg Med. 2026;25: 502-10

Calibration of Clinical Prediction Models: A Practical Guide to Interpretation, Reporting, and Recalibration

Hussein Alfakeer[1], Bahadır Taşlıdere[1], Lukasz Szarpak[2], Ayman El-Menyar[3], Başar Cander[1]

[1]Bezmialem Vakıf University Faculty of Medicine, Department of Emergency Medicine, İstanbul, Türkiye
[2]The John Paul II Catholic University of Lublin, Institute of Medical Science, Lublin, Poland
[3]Hamad Medical Corporation & Weill Cornell Medicine, Clinical Research and Trauma Surgery, Doha, Qatar

Abstract

Clinical prediction models are often evaluated primarily by discrimination, although many clinical applications depend on the accuracy of absolute risk estimates. Calibration describes agreement between predicted probabilities and observed event frequencies and must be assessed separately from discrimination. This narrative methodological review focuses on models that predict binary outcomes and examines how calibration should be assessed, interpreted, reported, and updated. The calibration hierarchy and the use of flexible calibration curves are described together with calibration-in-the-large, calibration slope, the integrated calibration index, E-statistics, and the Brier score. The review explains why grouped goodness-of-fit procedures, including the Hosmer-Lemeshow test, cannot confirm adequate calibration and why performance estimates require measures of uncertainty and a validation sample sized for the desired precision. Calibration of regression and machine learning models is considered without assuming that any model class is inherently well calibrated. Methods for intercept-only updating, logistic recalibration, and non-linear post-hoc calibration are presented, with attention to evaluating updated models in data not used for the updating step. Worked examples show how the direction, magnitude, and location of miscalibration affect interpretation. Because calibration alone does not establish clinical usefulness, models intended to guide treatment should also be evaluated for clinical utility at relevant decision thresholds. A reporting checklist specifies the information needed to judge whether risk estimates are suitable for the intended population and clinical context.

Keywords: Calibration, clinical prediction model, binary outcome, model validation, recalibration, clinical utility

Introduction

A clinical prediction model can rank patients correctly while producing inaccurate probabilities. An area under the receiver operating characteristic curve of 0.90 indicates strong discrimination, but it does not demonstrate that an estimated risk of 30% is numerically accurate. If only 12% of patients with predictions near 30% experience the outcome, the model overpredicts the risk in that part of the risk distribution. This error may change treatment recommendations when decisions are linked to risk thresholds. Discrimination and calibration therefore describe different properties of a prediction model (1).

Nevertheless, calibration is mentioned significantly less frequently than discrimination and is often misunderstood (2). Current resources explain the methods for assessing calibration and constructing calibration models (3), while reporting frameworks outline the necessary elements to report: the TRIPOD+AI statement now includes guidelines for reporting both regression and machine-learning prediction models (4). This guide is intentionally designed to supplement, rather than replicate, that guidance. While TRIPOD+AI addresses "what should be reported?" this article focuses on "how should the reported calibration be interpreted, and what actions should be taken?" It is structured as a workflow encompassing calibration, interpretation, reporting, and recalibration and is intended for repeated use when documenting or evaluating a prediction-model study.

Corresponding Author: Hussein Alfakeer MD, Bezmialem Vakıf University Faculty of Medicine, Department of Emergency Medicine, İstanbul, Türkiye
E-mail: husseinalfakeer1@gmail.com **ORCID ID:** orcid.org/0000-0003-1650-1160

Cite this article as: Alfakeer H, Taşlıdere B, Szarpak L, El-Menyar A, Cander B. Calibration of clinical prediction models: a practical guide to interpretation, reporting, and recalibration. Eurasian J Emerg Med. 2026;25: 502-10

Received: 02.09.2026
Accepted: 08.10.2026
Published: 06.11.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Scope and Review Approach

This article is a narrative methodological review of calibration methods for models that estimate the probability of a binary outcome. We considered methodological studies, reporting guidance, and applied tutorials addressing calibration assessment, validation, and model updating. Relevant sources were identified through PubMed and Google Scholar searches and by hand-searching the reference lists of key methodological papers and reporting guidelines. The article is intended to explain the interpretation and reporting of established methods. It is not a systematic review, and no formal assessment of risk of bias or certainty of evidence was undertaken. Calibration of time-to-event, competing-risk, ordinal, and multiclass predictions requires additional methods and is outside the present scope.

Discrimination and Calibration: Why AUC is not Enough

Discrimination and calibration are distinct concepts and largely independent (Table 1). Discrimination refers to a relative characteristic that relies solely on the ordering of predictions: the area under the curve (AUC) remains constant despite any transformation that maintains the order. Thus, even if every predicted risk is halved, the AUC stays the same, although calibration is compromised. Calibration, on the other hand, is an absolute characteristic based on precisely predicted values. Consequently, a high AUC does not guarantee that the predicted risks are numerically accurate (1). Figure 1 illustrates this: two models with an identical AUC of 0.85 can exhibit markedly different calibration levels-one may align closely with the observed risk, while the other may consistently overestimate it. Only the accurately calibrated model can be relied upon to generate risk estimates suitable for decision-making.

The Calibration Hierarchy

Calibration can be evaluated with varying degrees of rigor (5). Mean calibration, also known as calibration-in-the-large, requires that the average predicted risk aligns with the observed event rate. Weak calibration ensures that predictions are not consistently overestimated or underestimated (intercept 0) and are not excessively extreme or overly moderate (slope 1). Moderate calibration necessitates that predicted risks correspond to observed frequencies throughout the entire range, as indicated by a calibration curve that closely follows the diagonal. Strong calibration, which requires accurate probabilities for every covariate pattern, is practically impossible to achieve. In practical terms, moderate calibration is a suitable goal during validation and should be assessed using a flexible calibration curve in addition to metrics of weak calibration.

Reading a Calibration Plot

The primary instrument is the calibration plot, as depicted in Figure 2. The horizontal axis represents predicted probability, while the vertical axis shows observed frequency, with a 45° line indicating perfect calibration. Instead of categorizing patients into deciles, which can be noisy and dependent on cut-points (6), observed frequency should be estimated using a flexible smoother like loess or splines. Additionally, it's important to display the distribution of predicted risks alongside, as the curve is least reliable in areas with sparse predictions (1). The shape of the curve provides direct insight: if it falls below the diagonal, it indicates overprediction (Figure 2b); if it rises above the diagonal, it indicates underprediction (Figure 2c). A curve that is flatter than the diagonal, intersects it, and overestimates high risks while underestimating low risks signals overfitting; this is characterized by a calibration slope less than 1 (Figure 2d).

Numerical Calibration Metrics: How to Interpret Them

In addition to the calibration plot, several numerical measures were used to assess model calibration (Table 2). The calibration intercept, also referred to as calibration-in-the-large, measures the systematic difference between predicted and observed risks; its ideal value is 0. It was estimated using a logistic regression model with the linear predictor included as an offset, thereby fixing the calibration slope at 1. With this approach, a positive intercept indicates that the model tends to underestimate risk, whereas a negative intercept indicates overestimation. This

Table 1. Discrimination and calibration answer different questions

Aspect	Discrimination	Calibration
Question	Does the model rank patients with and without the event correctly?	Do the predicted probabilities match the observed frequencies?
Nature	Relative-depends only on ranking	Absolute-depends on the exact predicted values
Typical metric	C-statistic/AUC	Calibration plot, slope, intercept, ICI
Behavior	Unchanged by monotone rescaling of predictions	Can be altered by rescaling of predicted probabilities.
Blind spot	Says nothing about whether risks are numerically right	—
Practical limitation	—	More fragile on transport; needs an adequate sample

AUC: Area under the curve, ICI: Integrated calibration index

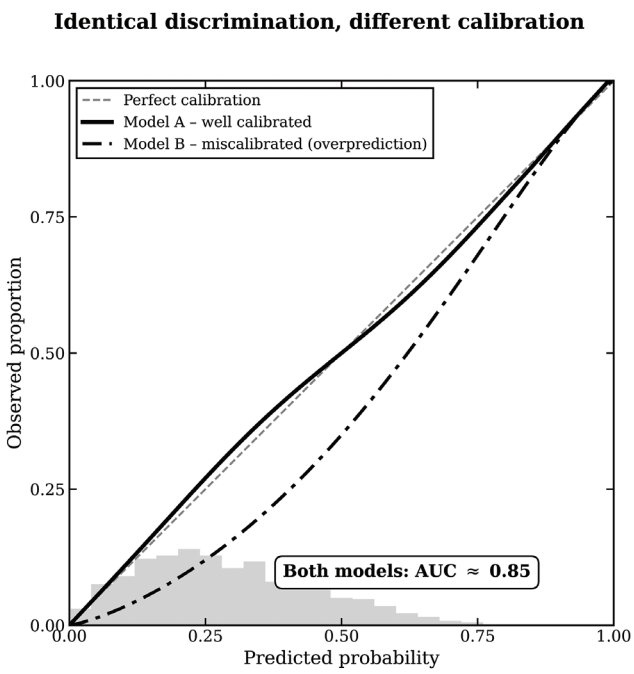


Figure 1. Schematic illustration of two models with the same discrimination and different calibration. The solid curve represents a model whose predicted probabilities remain close to the observed event frequencies. The dash-dot curve lies below the ideal line across most of the risk range, indicating systematic overprediction. The dashed diagonal represents perfect calibration. The shaded histogram shows an illustrative distribution of predicted probabilities. The curves are schematic and were not estimated from patient data

AUC: Area under the curve

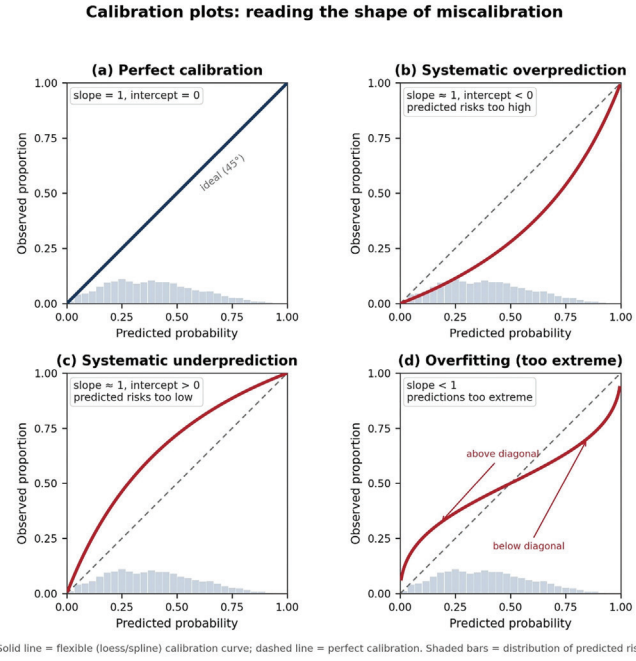


Figure 2. Four canonical calibration patterns. Schematic curves illustrating common calibration patterns; these curves are not derived from patient data. Predicted probabilities are shown on the horizontal axis and observed proportions on the vertical axis. The dashed line represents perfect calibration. The calibration intercept (calibration-in-the-large) is estimated with the slope fixed at 1. (a) Perfect calibration (slope =1, intercept =0). (b) Systematic overprediction, with the curve below the diagonal (intercept <0). (c) Systematic underprediction, with the curve above the diagonal (intercept >0). (d) Overfitting is characterized by a flatter curve that crosses the diagonal, lying above it at low predicted risks and below it at high predicted risks (slope <1). Each panel illustrates how to interpret the calibration pattern and its implications for reporting and recalibration (see Table 2 and the recalibration workflow)

differs from the intercept obtained when the calibration intercept and slope are estimated simultaneously in a logistic recalibration model [$\text{logit}(p) = \alpha + \beta \cdot LP$], because the intercept depends on the estimated slope and cannot be interpreted in the same way.

The calibration slope reflects how closely the predicted risks correspond to the observed outcomes, with an ideal value of 1. It is estimated as the coefficient of the linear predictor in the joint recalibration model. A slope below 1 suggests that predictions are too extreme, often as a result of overfitting, whereas a slope above 1 indicates that predictions are less extreme than they should be (5).

The integrated calibration index (ICI) and E-statistics (E50, E90, and Emax) were used to quantify the differences between the calibration curve and the ideal diagonal line. These measures represent, respectively, the mean, median, 90th percentile, and maximum of the absolute differences. The ICI is particularly useful for interpreting the magnitude of calibration error: for example, an ICI of 0.05 indicates an average difference of approximately five percentage points between predicted and observed risks (7).

The Brier score is a measure of overall prediction accuracy. It accounts for both discrimination and calibration and should therefore not be interpreted as a calibration measure alone (8). Table 2 summarizes the definitions and interpretations of these metrics, while the numerical values in the worked examples are illustrative.

The numerical values of calibration measures depend on the population in which they are estimated. The unscaled Brier score is influenced by outcome frequency, which limits comparisons across datasets with different event rates. Reporting the Brier score of a null model or the scaled Brier score can help quantify improvement over predictions based solely on the observed event rate. The ICI is weighted by the empirical distribution of predicted risks. A low ICI may therefore coexist with clinically relevant miscalibration in a narrow or sparsely represented part of the risk range. Emax should be interpreted cautiously because a small number of observations can determine its value.

It is important to exercise caution with thresholds: these metrics lack universal clinical cut-offs. Assigning a fixed value to determine whether an ICI, slope, or intercept is "good" or "bad" is risky, as the key factor is the extent of miscalibration in relation to the specific clinical decision being made. Metrics should be evaluated based on their inherent probability scale and the implications of overestimating or underestimating risk, rather than by comparing them to an arbitrary universal threshold (1).

Calibration is not a Single Number

Calibration cannot be encapsulated by a single statistic or hypothesis test. A model might achieve a satisfactory Brier score but still be inadequately calibrated, as the Brier score also acknowledges discrimination (8). A solitary summary, like the ICI, can obscure whether miscalibration is consistent or concentrated in a clinically significant segment of the risk spectrum. Thus, a comprehensive evaluation should include a minimum calibration set: (i) a flexible calibration plot showing the distribution of predicted risks, (ii) the calibration intercept, (iii) the calibration slope, (iv) an absolute measure such as the ICI or an E-statistic, and (v) the context in which these were validated (2). When considered together, these elements reveal both the direction and location of any miscalibration; however, when viewed separately, each can be misleading.

Uncertainty and Sample Size

Calibration statistics are sample estimates and should be reported with measures of precision. Confidence intervals should accompany the calibration intercept and slope. A flexible calibration curve should include pointwise confidence bands when these can be estimated reliably. The ICI and E-statistics are also subject to sampling uncertainty. Emax is particularly sensitive to sparse observations at the extremes of the predicted-risk distribution. Bootstrap methods can be used to obtain confidence intervals for these quantities.

The sample size for external validation should be determined by the precision required for the performance measures of interest. A minimum number of participants or outcome events alone does not ensure a reliable calibration analysis. The required sample size depends on outcome frequency, the distribution of predicted risks, anticipated model performance, and the desired precision of the calibration intercept, slope, and observed-to-expected ratio (9).

Calibration and Clinical Utility

Calibration is necessary when predicted risks are used to guide treatment, but calibration alone does not establish clinical usefulness. At a given threshold probability, models with similar calibration may identify different numbers of true-positive and false-positive cases. Model selection for clinical use should therefore include an explicit assessment of the consequences of threshold-based decisions. Decision curve analysis expresses these consequences as net benefit and allows comparison with default strategies such as treating all patients or treating none (10,11).

Discrimination, calibration, and clinical utility answer different questions. A model may have acceptable calibration and still provide no clinical benefit at the intended threshold. Conversely, the effect of miscalibration on clinical decisions may vary across the range of threshold probabilities. Claims about treatment selection should be supported by an analysis performed at clinically relevant thresholds.

Why a Non-significant Hosmer-Lemeshow Test does not Mean Good Calibration

The Hosmer-Lemeshow test remains a common method for "confirming" calibration, with a p-value exceeding 0.05 frequently interpreted as proof of a model's good calibration. However, this interpretation is incorrect. The test divides patients into bins, typically by deciles, and compares the observed counts with

Table 2. How to interpret common calibration metrics

Metric	Ideal	What does it tell you?	If abnormal
Calibration intercept	0	Overall systematic error (calibration-in-the-large)	>0 → underprediction (risks too low); <0 → overprediction (risks too high)
Calibration slope	1	Whether predictions are too extreme or too moderate	<1 → predictions too extreme (overfitting); >1 → too moderate
ICI	0	Mean absolute calibration error (probability scale)	Higher = worse; read as average percentage-point discrepancy
E50	0	Median absolute calibration error	Higher = worse
E90	0	90th-percentile calibration error	Higher = worse; flags a poorly calibrated tail
Emax	0	Maximum calibration error	Higher = worse; worst-case discrepancy
Brier score	0	Overall probabilistic accuracy (discrimination + calibration)	Lower = better; not calibration-specific

ICI: Integrated calibration index, E50/E90/Emax: median, 90th-percentile and maximum absolute calibration error

the expected ones. It has three significant flaws: the grouping is arbitrary; its power is highly dependent on sample size, making it underpowered in small samples and almost always significant in very large ones; and it only provides a p-value without indicating direction, magnitude, or where miscalibration occurs along the risk spectrum (12). Consequently, a non-significant result with a moderate sample size can still be consistent with clinically significant miscalibration. Therefore, this test should not be the primary or sole method for assessing calibration; the minimum calibration set mentioned earlier is much more informative (1).

Calibration Across Validation Settings

Calibration should be interpreted based on the method of assessment (Table 3). Calibration that appears accurate on development data tends to be overly optimistic; internal validation methods, such as bootstrapping or cross-validation, adjust for this optimism and often show a calibration slope less than 1. The most revealing test is external validation, conducted in a different time frame or context, where models frequently fail due to variations in case-mix and outcome prevalence affecting calibration more than discrimination (1). Since performance is tied to the specific population and timing of assessment, no model is permanently "validated" (13).

Calibration in Machine Learning Models: What Changes?

Calibration behavior cannot be inferred from the name of an algorithm alone. It depends on the loss function, regularization, hyperparameter tuning, class sampling, sample size, and the population in which the model is evaluated. Logistic regression fitted by maximum likelihood has, in its development data under standard conditions, an apparent calibration intercept of 0 and a slope of 1. This property does not guarantee that the flexible calibration curve follows the diagonal, nor does it extend to new populations. Penalization, class weighting, and sampling procedures may also alter apparent calibration.

Machine learning models may produce predictions that are too extreme, too moderate, or poorly calibrated otherwise. These patterns should be assessed using validation data rather than assumed based on the type of model (Table 4). A large systematic review found no consistent advantage of machine learning over logistic regression for clinical prediction models using structured data (14). However, findings vary across studies. Depending on the clinical setting, sample size, outcome, and modeling approach, machine learning models may perform similarly to or better than logistic regression (15).

Recalibration should be considered when validation reveals miscalibration large enough to affect the model's intended clinical use. Platt scaling uses logistic regression to adjust the original prediction scores, whereas isotonic regression allows a more flexible, non-linear adjustment but generally requires a larger calibration sample. Regardless of the method, calibration must be performed without using the outcomes from the final evaluation dataset. When the sample size is too small to establish separate training, calibration, and evaluation datasets, nested resampling provides an alternative for assessing model performance while accounting for the calibration process (16-18).

Model Recalibration

A stepwise workflow for assessing and managing calibration, from the initial assessment of discrimination and calibration to intercept updating, slope recalibration, and model revision, is summarized in Figure 3. Let p denote the probability produced by the original model. Intercept-only updating estimates a new intercept while fixing the coefficient of the original linear predictor at 1:

$\text{logit}(p^*) = a + \text{logit}(p)$.

This update corrects calibration-in-the-large while preserving the relative spread of the original predictions. A negative value of a is expected when the original model systematically overpredicts risk.

Logistic recalibration estimates both an intercept and a slope:

$\text{logit}(p^*) = a + b\ \text{logit}(p)$.

A value of b below 1 shrinks predictions that are too extreme. A value above 1 increases the spread of predictions that are overly moderate. This model can correct weak calibration, but cannot remove non-linear deviations of the flexible calibration curve from the diagonal. More flexible methods, including isotonic regression, may address non-linear miscalibration but can overfit small calibration datasets.

Table 3. Validation settings and what each says about calibration

Setting	What is it telling us?	Main limitation
Apparent (development data)	In-sample fit; upper bound on performance	Optimistic; ignores overfitting
Bootstrap	Internal optimism correction	Same source population
Cross-validation	Internal validation	May not capture transportability
Temporal validation	Performance over time; drift	Setting-specific
Geographic/external validation	Transportability to new settings	Requires an external cohort

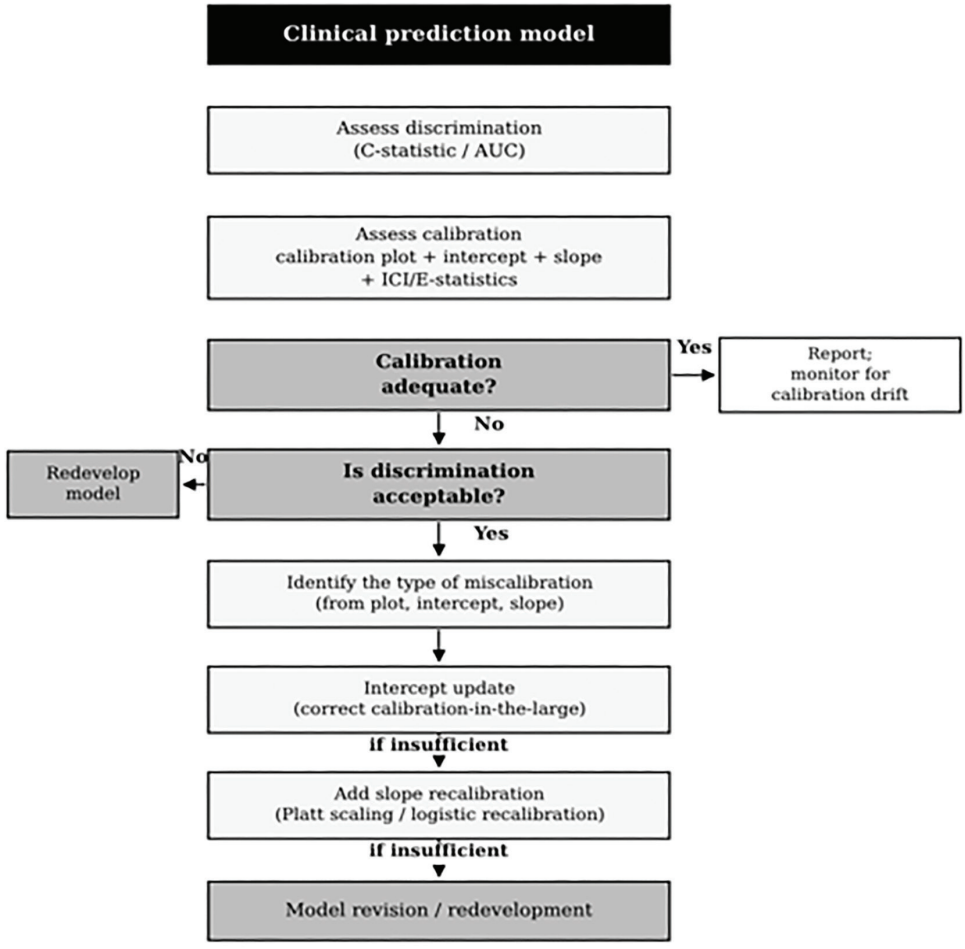


Figure 3. A workflow for assessing and managing calibration. Discrimination and calibration are assessed first; if calibration is inadequate but discrimination is acceptable, the type of miscalibration is identified and corrected through escalating steps: intercept update, slope recalibration, and model revision.

AUC: Area under the curve, ICI: Integrated calibration index

Table 4. Calibration behavior of common model classes and the strategy usually used to obtain calibrated probabilities

Model	Output probabilities	Common concern in validation	Typical strategy
Logistic regression	Yes	Optimism in development data; non-linear miscalibration; changes in baseline risk or predictor effects during transport.	Validate calibration in the intended setting; update the intercept or slope only when indicated
Random forest	Yes (vote proportions)	Model-dependent miscalibration; compressed or irregular probability estimates may occur.	Platt scaling or isotonic regression
Gradient boosting	Yes	Model-dependent overconfidence or underconfidence.	Platt scaling or isotonic regression
Support vector machine	No (decision scores)	Decision scores are not probabilities and require a separate probability mapping.	Platt scaling to obtain probabilities
Neural network	Yes (sigmoid/softmax)	Overconfidence may occur, particularly after extensive fitting.	Temperature scaling, Platt scaling, or isotonic regression

Recalibration should be treated as a model-updating procedure. The performance reported in the data used to estimate a and b is optimistic. The updated model should be evaluated on new data or by a resampling procedure that repeats the entire updating process. The final report should provide the updated equation, the source and size of the recalibration sample, and calibration and clinical-utility results after updating (18,19).

Worked Examples

The following examples illustrate how the metrics are interpreted together. Values are illustrative.

Example 1) Strong Discrimination, Inadequate Calibration: A model presents an AUC of 0.91, a calibration intercept of -0.42, a slope of 0.63, and an ICI of 0.071. Although discrimination is strong, the negative intercept indicates an overall overprediction, and the estimated slope of 0.63 suggests that the predictions are too extreme. The ICI indicates that predicted and actual risks differ, on average, by approximately seven percentage points. Even with the high AUC, these absolute risks should not be used for decision-making without recalibration.

Example 2) Moderate Discrimination, Good Calibration: A model indicates an AUC of 0.78, an intercept of 0.03, a slope of 0.97, and an ICI of 0.021. Although discrimination is only moderate, the intercept is close to 0, the slope is near 1, and the average calibration error is roughly two percentage points, which together demonstrate that the predicted probabilities are closely aligned with the observed risk. These estimates are consistent with moderate discrimination and mild miscalibration in this sample. They do not establish that the model should be used to determine treatment at a 10% threshold. That decision requires comparing net benefit at 10% and considering the relative consequences of false-positive and false-negative classifications. Neither AUC nor calibration metrics provide this information on their own.

Example 3) External Validation: During the development phase, a model achieves an AUC of 0.88 with a slope of 1.00 and an intercept of 0.00. However, in external validation, the AUC decreases to 0.84, with a slope of 0.71, an intercept of -0.31, and an ICI of 0.058. While discrimination remains largely preserved, calibration has worsened: the slope has dropped to 0.71, indicating that predictions are too extreme in the new context; the negative intercept suggests overprediction. The negative intercept is consistent with a lower event rate in the validation setting, but the performance estimates do not identify the cause of miscalibration. Differences in case mix, predictor measurement, clinical practice, outcome definition, or predictor effects should also be considered. This pattern is common in transport scenarios, where calibration issues arise before discrimination problems, necessitating recalibration (updating the intercept and likely adjusting the slope) before applying the model locally. Published examples of calibration findings in clinical prediction models are summarized in Table 5.

A Ten-point Calibration Reporting Checklist

- Is the predicted outcome and, where relevant, the prediction horizon defined?
- Is the population, setting, and type of validation stated?
- Is a flexible calibration curve presented together with the distribution of predicted risks?
- Are confidence intervals or confidence bands reported for the main calibration estimates?
- Are calibration-in-the-large and the calibration slope reported with their estimation methods?
- Is an absolute calibration measure reported and interpreted on the probability scale?
- Is the sample size justified according to the precision required for calibration assessment?
- Is calibration examined over the clinically relevant risk range and in important subgroups when supported by the sample size?
- Is clinical utility evaluated at prespecified decision thresholds when the model was intended to guide an intervention?
- If the model is updated, are the updating equation, data used for updating, and subsequent performance reported?

Conclusion

Calibration describes whether predicted probabilities agree with observed risks in a defined population and time period. It should be assessed with a flexible calibration curve, calibration-in-the-large, a calibration slope, and measures of uncertainty. The ICI, E-statistics, and Brier score provide supplementary information, but their interpretation depends on sample size, outcome frequency, and the distribution of predicted risks.

Miscalibration may be corrected by intercept updating or logistic recalibration when the original model retains useful predictive information. Any updated model requires further evaluation. When predictions are intended to guide treatment, calibration results should be considered together with clinical utility at the relevant decision thresholds. Performance should be reassessed after being transported to another setting and when changes in outcome incidence, predictor measurements, or clinical practice are suspected.

Table 5. Examples of calibration in clinical prediction models

Clinical prediction model	Outcome and validation setting	Reported performance	Calibration finding	Practical interpretation
QRISK3	Ten-year cardiovascular disease risk; independent external validation in 403,370 UK Biobank participants	Harrell C-statistic 0.722 in women and 0.697 in men	Systematic overprediction, particularly in older participants, by as much as 20%	Moderate discrimination did not ensure accurate absolute risk; recalibration or an alternative model may be needed for UK Biobank applications (20).
Spanish MBDS ischemic-stroke mortality model	In-hospital mortality after ischemic stroke; temporal external validation in a 2016-2018 Spanish national cohort	AUC 0.726 on external validation and 0.743 after coefficient refitting	All model coefficients were re-estimated in the later cohort, modestly improving discrimination; this was model refitting rather than simple intercept or slope recalibration	The study illustrates temporal model drift and coefficient refitting before reuse in a later population; comprehensive calibration measures were not reported (21).
Cardiovascular clinical prediction models, pooled literature	Multiple cardiovascular outcomes; 2,030 external validations of 1,382 models	Median external-validation AUC 0.73; 81% of validations reporting AUC showed lower discrimination than in development data	Only 53% of validations reported any calibration measure, and performance varied substantially across settings	External validation commonly reveals performance loss, while incomplete calibration reporting limits assessment of whether absolute risks are clinically usable (22).
ISARIC 4C Deterioration model	In-hospital clinical deterioration among adults hospitalized with COVID-19; held-out validation in 8,239 patients from London	C-statistic 0.77 (95% CI 0.76-0.78)	Calibration-in-the-large 0.00 (95% CI -0.05 to 0.05) and calibration slope 0.96 (95% CI 0.91-1.01)	Moderate discrimination was accompanied by near-ideal weak calibration in the held-out region, illustrating that calibration and discrimination should be reported jointly (23).
EuroSCORE II	Thirty-day mortality after cardiac surgery for infective endocarditis; nationwide Netherlands registry, 2,569 patients	AUC 0.72; O: E ratio 0.69	Overall mortality was overpredicted, with overprediction concentrated above a 20% predicted probability	Acceptable ranking can coexist with risk-range-specific miscalibration; high predicted risks require caution, and performance was poor in emergency surgery (24).
PREDICT Breast v3.1	Ten-year overall survival after early breast cancer; external validation in national cohorts from the Netherlands and Sweden	AUC 0.75 in both national cohorts	Calibration was near ideal overall in Sweden and showed slight underestimation of mortality in the Netherlands; slopes were lower in patients aged ≤40 years	Good average calibration at the population level can mask clinically relevant subgroup miscalibration, supporting subgroup-specific assessment (25).

Findings are derived from different validation studies and are not directly comparable across models, outcomes, populations, or study settings.
AUC: Area under the curve, CI: Confidence interval, ISARIC: International severe acute respiratory and emerging infections consortium, MBDS: Minimum basic data set, O:E ratio: Observed-to-expected events ratio.

Footnotes

Author Contributions

Concept: H.A., B.T., L.S., A.E.-M., B.C., Design: H.A., B.T., L.S., A.E.-M., B.C., Literature Search: H.A., B.T., L.S., A.E.-M., B.C., Writing: H.A., B.T., L.S., A.E.-M., B.C.

Conflict of Interest: Bahadır Taşlıdere, Lukasz Szarpak, and Başar Cander are members of the Editorial Board of the Eurasian Journal of Emergency Medicine. However, they were not involved in any stage of the editorial decision process for this manuscript. The editors who evaluated this manuscript are from different institutions. The other authors declared no conflict of interest.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW; Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17:230.
2. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. JAMA. 2017;318:1377-84.
3. Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. J Am Med Inform Assoc. 2020;27:621-33.

4. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378.
5. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol. 2016;74:167-76.
6. Van Hoorde K, Van Huffel S, Timmerman D, Bourne T, Van Calster B. A spline-based tool to assess and visualize the calibration of multiclass risk predictions. J Biomed Inform. 2015;54:283-93.
7. Austin PC, Steyerberg EW. The integrated calibration index (ICI) and related metrics for quantifying the calibration of logistic regression models. Stat Med. 2019;38:4051-65.
8. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21:128-38.
9. Riley RD, Snell KIE, Archer L, Ensor J, Debray TPA, van Calster B, et al. Evaluation of clinical prediction models (part 3): calculating the sample size required for an external validation study. BMJ. 2024;384:e074821.
10. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. Med Decis Making. 2015;35:162-9.
11. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26:565-74.
12. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Crit Care Med. 2007;35:2052-6.
13. Van Calster B, Steyerberg EW, Wynants L, van Smeden M. There is no such thing as a validated prediction model. BMC Med. 2023;21:70.
14. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12-22.
15. İkiz F, Ak A. Machine learning vs. conventional statistics in predicting medical outcomes: a comparative performance analysis using a viral disease dataset. Eurasian J Emerg Med. 2026;25:228-36.
16. Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. In: proceedings of the 34th International Conference on Machine Learning (ICML). PMLR. 2017;70:1321-30.
17. Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. In: Proceedings of the 22nd International Conference on Machine Learning (ICML). 2005:625-32.
18. Jiang X, Osl M, Kim J, Ohno-Machado L. Calibrating predictive model estimates to support personalized medicine. J Am Med Inform Assoc. 2012;19:263-74.
19. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. 2nd ed. Cham: Springer; 2019.
20. Parsons RE, Liu X, Collister JA, Clifton DA, Cairns BJ, Clifton L. Independent external validation of the QRISK3 cardiovascular disease risk prediction model using UK Biobank. Heart. 2023;109:1690-7.
21. García-Torrecillas JM, Lea-Pereira MC, Amaya-Pascasio L, Rosa-Garrido C, Quesada-López M, Reche-Lorite F, et al. External validation and recalibration of a mortality prediction model for patients with ischaemic stroke. J Clin Med. 2023;12:7168.
22. Wessler BS, Nelson J, Park JG, McGinnes H, Gulati G, Brazil R, et al. External validations of cardiovascular clinical prediction models: a large-scale review of the literature. Circ Cardiovasc Qual Outcomes. 2021;14:e007858.
23. Gupta RK, Harrison EM, Ho A, Docherty AB, Knight SR, van Smeden M, et al. Development and validation of the ISARIC 4C deterioration model for adults hospitalised with COVID-19: a prospective cohort study. Lancet Respir Med. 2021;9:349-59.
24. Heinen FJ, Peijster AJL, Fu EL, Kamp O, Chamuleau SAJ, Post MC, et al. External validation of EuroSCORE I and II in patients with infective endocarditis: results from a nationwide prospective registry. Eur J Cardiothorac Surg. 2024;66:ezae418.
25. Verheul EM, Doornkamp F, Petrov I, Siesling S, Lingsma HF, Koppert LB, et al. External validation of PREDICT Breast v3.1 for overall survival in international cohorts, including young and invasive lobular subgroups. Breast Cancer Res Treat. 2026;217:14.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.40360
Eurasian J Emerg Med. 2026;25: 511-9

Cecal Volvulus Within the Etiological Spectrum of Intestinal Obstruction in Adults Admitted to General Surgery: A Single-center Retrospective Cohort Study

Serkan Demir[1]*, Muhammed Salih Süer[1], Ahmet Yiğit Kalelioğlu[1], Faruk Yazıcı[2], Engin Ölçücüoğlu[1], Bahattin Bayar[1], Mert Şahin[3], Hacer Demirdelen[4], Lukasz Szarpak[5,6]

[1]University of Health Sciences Türkiye, Ankara Etlik City Hospital, Clinic of General Surgery, Ankara, Türkiye
[2]Etimesgut Şehit Sait Ertürk State Hospital, Clinic of General Surgery, Ankara, Türkiye
[3]Pursaklar Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[4]University of Health Sciences Türkiye, Ankara Etlik City Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[5]The John Paul II Catholic University of Lublin, Collegium Medicum, Institute of Medical Science, Lublin, Poland
[6]Baylor College of Medicine, Henry J.N. Taub Department of Emergency Medicine, Houston, USA
*Current affiliation: Yerköy Şehit Korgeneral Osman Erbaş State Hospital, Clinic of General Surgery, Yozgat, Türkiye

Abstract

Aim: To describe the etiological distribution of acute intestinal obstruction among adults admitted to general surgery, the anatomical distribution of volvulus, and the characteristics of the cecal volvulus subgroup.

Materials and Methods: Adults admitted to the General Surgery Clinic of University of Health Sciences Türkiye, Ankara Etlik City Hospital, via the emergency department or by direct hospitalization, with acute intestinal obstruction or ileus between October 2022 and October 2025 were retrospectively analyzed. The cohort was classified by etiology, and volvulus cases were categorized as cecal or non-cecal. The cecal volvulus subgroup was analyzed descriptively, without inferential statistics.

Results: Among 1.316 patients (696 male; mean age 62.7±18.3 years), adhesive ileus was the most common etiology (861, 65.4%), and volvulus was identified in 109 patients (8.3%): 105 non-cecal and 4 cecal. The four cecal volvulus patients were aged 22-83 years; three had a neurological comorbidity. The original computed tomography (CT) report identified cecal volvulus in two patients. One preoperative endoscopic detorsion attempt was unsuccessful. All four underwent open resection; the two with ischemia and necrosis received a stoma, and the other two received a primary anastomosis. One 83-year-old patient died within 90 days from postoperative cardiopulmonary complications.

Conclusion: In this cohort of surgically admitted patients with intestinal obstruction, cecal volvulus was uncommon. The descriptive findings in this four-patient subgroup are hypothesis-generating and consistent with recommendations in the literature that the anatomical site of volvulus be explicitly reported on CT.

Keywords: Cecal volvulus, intestinal obstruction, emergency surgery, computed tomography, sigmoid volvulus, diagnostic delay

Corresponding Author: Serkan Demir MD, University of Health Sciences Türkiye, Ankara Etlik City Hospital, Clinic of General Surgery, Ankara, Türkiye
E-mail: mdserkandemir@gmail.com **ORCID ID:** orcid.org/0009-0008-1187-7883

Cite this article as: Demir S, Süer MS, Kalelioğlu AY, Yazıcı F, Ölçücüoğlu E, Bayar B, et al. Cecal volvulus within the etiological spectrum of intestinal obstruction in adults admitted to general surgery: a single-center retrospective cohort study. Eurasian J Emerg Med. 2026;25: 511-9

Received: 23.05.2026
Accepted: 19.10.2026
Published: 09.11.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

Introduction

Intestinal obstruction is a common surgical emergency requiring rapid clinical assessment and systematic etiological workup. Depending on whether the origin is mechanical or functional, the clinical spectrum varies considerably (1). Adhesive ileus, colorectal tumor obstruction, hernia, paralytic ileus, and volvulus are the principal etiologies, each requiring a distinct diagnostic and therapeutic approach (2). Accordingly, in the emergency setting, the diagnosis of "ileus" or "intestinal obstruction" should not be regarded as a definitive endpoint, but rather as a starting point from which the underlying etiology must be systematically investigated (1).

Volvulus is a rare but surgically critical cause of intestinal obstruction (3). When characterized by axial rotation around the mesenteric axis, it may rapidly progress to ischemia and perforation. Sigmoid volvulus is the most prevalent type, occurring particularly in elderly patients with chronic constipation and neurological comorbidities (4). In selected stable patients, sigmoid volvulus can be managed with endoscopic detorsion. Cecal volvulus, by contrast, responds poorly to endoscopic treatment and is usually managed surgically (3).

The clinical presentation of cecal volvulus may overlap with sigmoid volvulus and nonspecific obstruction, creating a risk of diagnostic delay or misclassification (3). Abdominal pain, distension, vomiting, and cessation of flatus are common to both conditions, complicating etiological differentiation (4). However, the diagnostically and therapeutically critical question is not merely "is volvulus present?" but also "where is the volvulus located?" given that the management algorithms for sigmoid and cecal volvulus differ fundamentally.

Computed tomography (CT) is the primary imaging modality for etiological differentiation of intestinal obstruction and for determining the anatomical localization of volvulus (5). Systematic evaluation of the whirl sign, bird-beak appearance, ectopic cecal position, transition point, and ischemic findings is essential to distinguish cecal volvulus from sigmoid volvulus or nonspecific ileus. However, missed or inadequate CT classification has been reported (6).

The aim of this study was twofold. First, we aimed to describe the etiological distribution of acute intestinal obstruction among adults admitted to the general surgery department of a tertiary care center over a three-year period and the anatomical distribution of volvulus within this cohort. Second, we aimed to describe the cecal volvulus subgroup identified within this cohort in detail, providing complete patient-level clinical, radiological, operative, and outcome data. Because only four patients had cecal volvulus, this subgroup analysis was not designed to test hypotheses or estimate diagnostic performance, risk factors, or treatment effects; its findings are descriptive and hypothesis-generating, and are interpreted in the context of the existing literature.

Materials and Methods

Study Design and Ethics

This single-center retrospective cohort study was based on the archival records of the General Surgery Clinic of University of Health Sciences Türkiye, Ankara Etlik City Hospital and included 1.316 adults admitted to that department with intestinal obstruction. The primary analysis described the etiological distribution within the cohort and the anatomical distribution of volvulus; the secondary analysis provided a detailed evaluation of the cecal volvulus subgroup (n=4) identified within the cohort. No sample size or power calculation was performed because the cecal volvulus subgroup was too small for statistical inference. The study was reported in accordance with the STROBE guidelines. The study was approved by the Clinical Research Ethics Committee of University of Health Sciences Türkiye, Ankara Etlik City Hospital (decision number: AEŞH-BADEK2-2025-745, date: 30.12.2025) and conducted in accordance with the Declaration of Helsinki.

Patient Selection and Inclusion/Exclusion Criteria

Adult patients (≥18 years) admitted to the General Surgery Department between October 2022 and October 2025 with a diagnosis of intestinal obstruction or ileus, established on the basis of clinical, laboratory, and imaging findings, were retrospectively screened. Patients were admitted via the emergency department or by direct hospitalization. Because only patients admitted to the general surgery department were included, the cohort does not represent all patients who presented to the emergency department with obstruction; those managed by other departments, discharged from the emergency department, or transferred to other centers were not captured.

The following patients were excluded from the study:

- Pediatric patients under 18 years of age
- Patients with insufficient or incomplete archival records
- Patients in whom the diagnosis of obstruction could not be clinically confirmed
- Patients who underwent surgery at an external center and were subsequently transferred to our department

Etiological Classification

Included patients were assigned to one of the following mutually exclusive etiological categories on the basis of clinical, radiological, and operative findings:

- Adhesive ileus (mechanical obstruction secondary to adhesions)
- Colonic tumor obstruction
- Small bowel mass obstruction
- Volvulus
- Other and unidentifiable causes (including intussusception and bezoar)

No patient in the final cohort was assigned to a hernia-related category. Patients with obstruction caused by an incarcerated or strangulated external abdominal wall hernia were recorded under a separate hernia diagnosis in our departmental archive and were, therefore, not included in this obstruction cohort.

Volvulus Subgroup Analysis

Patients with volvulus were classified by anatomical location as cecal or non-cecal volvulus. The non-cecal group (sigmoid and other colonic volvulus) was not further subclassified by anatomical site. The diagnosis of cecal volvulus was confirmed by a combined assessment of CT findings, operative notes, and pathological results.

Data Collection

Demographic data, comorbidity profile, presenting symptoms, vital signs, laboratory parameters, imaging findings, operative findings, postoperative course, and outcomes were extracted from archival records. For cecal volvulus patients, original CT reports were recorded, and the responsible radiologist performed a secondary radiological re-evaluation. As this re-evaluation was non-blinded and conducted with awareness of the surgical diagnosis, a risk of confirmation bias was acknowledged. Detailed patient-level variables were analyzed for the four cecal volvulus patients; for the remaining patients, only the data required for etiological classification and demographic description were used.

Statistical Analysis

Data were analyzed using SPSS version 26.0 (IBM Corp., Armonk, NY, USA). For the overall cohort, continuous variables are expressed as mean ± standard deviation and categorical variables as frequencies and percentages. For the four patients with cecal volvulus, data are presented individually (Table 1) and summarized as counts (n/N) and medians (range) (Table 2); percentages are not reported for this subgroup, and no inferential statistical tests were performed. Ninety-day all-cause mortality was ascertained through routine outpatient follow-up and cross-referenced with national electronic health records. CT findings are reported as case-level observations, rather than as measures of diagnostic accuracy.

Results

General Cohort

A total of 1,316 patients who were admitted to our department with acute intestinal obstruction or ileus between October 2022 and October 2025 met the inclusion criteria. Of these, 696 (52.9%) were male and 620 (47.1%) were female, with a mean age of 62.7±18.3 years.

Regarding the etiological distribution, 861 patients (65.4%) presented with adhesive ileus, 297 (22.6%) with colonic tumor obstruction, 27 (2.1%) with small bowel mass obstruction, 109 (8.3%) with volvulus, and 22 (1.7%) with other or unidentifiable causes. Of the 22 patients in this last group, 18 were evaluated as having intussusception-related obstruction and 4 as having bezoar-related obstruction.

Volvulus Subgroup

Among the 109 patients with volvulus, 4 (3.7%) had cecal volvulus and 105 (96.3%) had non-cecal (sigmoid or other colonic) volvulus; the non-cecal group was not further subclassified by anatomical site. No comparative analysis between the cecal and non-cecal groups was performed.

Cecal Volvulus Subgroup

Individual patient data are presented in Table 1 and summarized in Table 2. The four patients (two males and two females) were aged 22, 30, 57, and 83 years (median 43.5 years). Three patients had a neurological comorbidity: two had intellectual disability and one had epilepsy with cerebral palsy; the 83-year-old patient had none. Three patients had undergone abdominal surgery.

All four patients presented with abdominal pain and cessation of flatus and stool; vomiting was present in three patients and abdominal distension in two patients. Symptom duration before admission ranged from 1 to 4 days. All patients were hemodynamically stable on admission, and peritoneal signs were present in one patient.

CT was performed in all four patients. The original CT reports specified cecal volvulus in two patients; in the other two (Patients 1 and 3), the reports described obstructive or nonspecific ileus; cecal volvulus was diagnosed intraoperatively. The whirl sign was reported in two patients. Endoscopic detorsion was attempted preoperatively in one patient (Patient 4) but was unsuccessful.

During surgery, a 360-degree rotation of the cecum was found in all four patients. Patients 1 and 2 had ischemia and necrosis, with gangrenous segments measuring 15 cm and 70 cm, respectively.

Table 1. Clinical, radiological, operative, and postoperative characteristics of patients with cecal volvulus (n=4)				
Variable	**Patient 1**	**Patient 2**	**Patient 3**	**Patient 4**
Demographics and comorbidities				
Age (years)	83	22	57	30
Sex	Male	Male	Female	Female
BMI (kg/m^2)	24.1	14.8	32.0	-
Neurological comorbidity	-	Epilepsy, cerebral palsy	Intellectual disability	Intellectual disability
Hypertension	No	No	Yes	No
Diabetes mellitus	No	No	No	No
Cardiac failure	No	No	No	No
COPD	No	No	No	No
Malignancy history	No	No	No	No
Prior abdominal surgery	Yes	Yes	Yes	No
Prior surgery type	Inguinal hernia repair	Gastric surgery	Nephrectomy	-
Anticoagulant use	No	No	No	No
Steroid use	No	No	No	No
Clinical presentation				
Symptom duration prior to admission (days)	3	2	1	4
Admission mode	**Referral (transfer)**	**Emergency**	**Emergency**	**Emergency**
Abdominal pain	Yes	Yes	Yes	Yes
Vomiting	Yes	No	Yes	Yes
Cessation of flatus and stool	Yes	Yes	Yes	Yes
Abdominal distension	Yes	No	Yes	No
Blood pressure (mmHg)	130/70	145/85	110/85	105/75
Heart rate (bpm)	93	82	110	105
Temperature (°C)	37.0	37.0	37.5	37.2
SpO_2 (%)	93	94	95	96
Hemodynamic status	Stable	Stable	Stable	Stable
Peritoneal signs	No	No	Yes	No
Laboratory findings at admission				
WBC ($\times10^3$/µL)	20.99	9.01	9.90	15.80
Neutrophil (%)	91.8	82.9	87.3	86.8
CRP (mg/L)	101.0	79.0	91.2	6.2
Lactate (mmol/L)	4.04	8.30	1.60	1.80
Sodium (mEq/L)	144	147	138	136
Potassium (mEq/L)	4.30	4.08	3.90	3.60
Chloride (mEq/L)	110	107	106	-
Creatinine (mg/dL)	1.03	0.40	1.10	0.50
pH	7.30	7.29	7.40	7.49
HCO_3 (mEq/L)	18.5	16.0	23.0	22.9
Base excess (mEq/L)	-7.3	-8.3	-3.9	-2.0
CT findings				
CT performed	Yes	Yes	Yes	Yes
Correct CT diagnosis of cecal volvulus	No*	Yes	No**	Yes
Whirl sign	No	Yes	No	Yes
Wall thickening	No	No	No	No
Free fluid on CT	No	Yes	Yes	Yes
Reduced contrast enhancement	No	No	No	No
Pneumoperitoneum	No	No	No	No
Mesenteric venous filling defect	No	No	No	No

Table 1. Continued

Variable	Patient 1	Patient 2	Patient 3	Patient 4
Operative findings and procedure				
Endoscopic detorsion attempted	No	No	No	Yes (failed)***
Cecal rotation (degrees)	360°	360°	360°	360°
Surgical approach	Open	Open	Open	Open
Bowel ischemia	Yes	Yes	No	No
Necrosis	Yes	Yes	No	No
Gangrenous segment length (cm)	15	70	-	-
Perforation	No	No	No	No
Intraoperative free fluid	No	Yes	No	Yes
Resection type	Ileocolic resection	Right hemicolectomy	Ileocolic resection	Ileocolic resection
Stoma formation	Yes (double-barrel)	Yes (end ileostomy)	No	No
Anastomosis type	-	-	Stapler	Stapler
Operation duration (min)	80	65	135	150
Postoperative course and outcomes				
ICU admission	Yes	Yes	Yes	Yes (brief observation)
ICU length of stay (days)	55	2	2	0
Postoperative intubation	Yes	No	No	No
Vasopressor requirement	Yes	No	No	No
First flatus (days)	1	3	4	5
First bowel movement (days)	1	3	5	8
Oral intake initiation (days)	Not resumed‡	3	6	4
Any complication	Yes	No	No	Yes
Clavien-Dindo grade	V****	-	-	IIIb†
Anastomotic leak	-	No	No	No
Surgical site infection	No	No	No	No
Postoperative ileus	No	No	No	No
Reoperation	No	No	No	Yes††
Hospital length of stay (days)	55	9	8	9
30-day mortality	No	No	No	No
90-day mortality	Yes	No	No	No
Recurrence of cecal volvulus	No	No	No	No
Stoma closure	No	No	N/A	N/A

*:Patient 1's CT report: "Diffuse calibration increase and air-fluid levels in the right mid and lower quadrant; max diameter 8 cm - obstructive ileus." Cecal volvulus confirmed intraoperatively.
**: Patient 3 CT report: "Dilated loop at midline level (~8 cm diameter) with air-fluid levels and collapsed distal loops - compatible with ileus. "Transition zone not identifiable due to absence of oral contrast." Cecal volvulus confirmed intraoperatively.
***: Patient 4: Endoscopic detorsion attempted preoperatively by gastroenterology; unsuccessful due to right-sided torsion anatomy. Emergency open surgery was performed subsequently.
****: Patient 1: Postoperative cardiopulmonary organ dysfunction requiring prolonged intensive care and progressing to death within 90 days (Clavien-Dindo Grade V; death attributable to these complications and advanced age).
†: Patient 4: Drain migration into the abdominal cavity was detected on postoperative day 6; an emergency laparotomy was performed for foreign body removal (Clavien-Dindo Grade IIIb).
††: Patient 4 reoperation: Drain removed via emergency laparotomy on postoperative day 6.
‡: Patient 1 did not resume oral intake; nutrition was provided enterally via a nasogastric tube from postoperative day 1 onward during the prolonged intensive care unit stay.
WBC: White blood cell count, CRP: C-reactive protein, CT: Computed tomography, ICU: Intensive care unit, COPD: Chronic obstructive pulmonary disease, BMI: body mass index

All patients underwent open surgical resection: ileocolic resection in three patients and right hemicolectomy in one patient. A stoma without an anastomosis was created in the two patients with ischemia and necrosis, and a stapled primary anastomosis was performed in the other two patients. Representative radiological and intraoperative findings are shown in Figure 1.

All four patients were admitted to the intensive care unit postoperatively for routine monitoring at the request of the

Table 2. Postoperative outcomes of cecal volvulus patients (n=4)

Outcome variable	Cecal volvulus patients (n=4)
Postoperative care	
Postoperative ICU admission (routine monitoring), n/N	4/4
ICU stay, median (range), days	2 (0-55)
Postoperative intubation, n/N	1/4
Vasopressor requirement, n/N	1/4
Gastrointestinal recovery	
First flatus, median (range), days	3.5 (1-5)
First bowel movement, median (range), days	4 (1-8)
Initiation of oral intake, median (range), days	4 (3-6)†
Complications	
Any complication, n/N	2/4
Highest Clavien-Dindo grade, n	IIIb: 1; V: 1*
Anastomotic leak, n/N	0/2
Surgical site infection, n/N	0/4
Postoperative ileus, n/N	0/4
Reoperation, n/N	1/4**
Hospital course	
Hospital length of stay, median (range), days	9 (8-55)
Excluding Patient 1 (prolonged ICU stay), median (range), days	9 (8-9)
Mortality and follow-up	
30-day mortality, n/N	0/4
90-day mortality, n/N	1/4
Recurrence of cecal volvulus, n/N	0/4
Stoma closure performed, n/N	0/2

*: Patient 4 experienced a Grade IIIb complication (reoperation for drain migration). Patient 1 died within 90 days due to postoperative cardiopulmonary complications (Clavien-Dindo Grade V).
**: Reoperation was performed in Patient 4 for drain migration (foreign-body removal via laparotomy). n/N: Number of patients with the outcome/number of patients at risk (anastomotic leak and stoma closure: two patients each), ICU: Intensive care unit.
†: Calculated for the three patients who resumed oral intake; Patient 1 did not resume oral intake

anesthesiology team; three were transferred to the ward within 48 hours, whereas Patient 1 required prolonged intensive care. Patient 1, an 83-year-old American Society of Anesthesiologists class 3 emergency (ASA 3E) transferred from another hospital after a 3-day symptom duration, with an admission lactate of 4.04 mmol/L and a 15-cm gangrenous segment, underwent stoma formation and died within 90 days from postoperative cardiopulmonary complications following prolonged intubation (Clavien-Dindo grade V). Patient 4 required reoperation on postoperative day 6 for intra-abdominal drain migration (Clavien-Dindo grade IIIb). The remaining two patients had no complications. Hospital length of stay ranged from 8 to 55 days (median 9 days); excluding Patient 1, it ranged from 8 to 9 days. No recurrence was observed.

Discussion

In this single-center cohort of 1.316 adults admitted to a general surgery department for acute intestinal obstruction or ileus over three years, adhesive ileus was the dominant etiology (861 patients, 65.4%), followed by colonic tumor obstruction (297, 22.6%), volvulus (109, 8.3%), small-bowel mass obstruction (27, 2.1%), and other or unidentifiable causes (22, 1.7%). Cecal volvulus was identified in four patients, all of whom underwent open resection. Because the study population was restricted to surgically admitted patients, these frequencies describe the case mix of a tertiary general surgery service rather than that of all emergency department presentations with obstruction. As emphasized in previous reviews, the diagnosis of "ileus" or "intestinal obstruction" should prompt systematic etiological clarification rather than be accepted as a final diagnosis (1,2,7).

The predominance of adhesive obstruction is consistent with the Bologna guidelines, which identify adhesions as the leading cause of mechanical small-bowel obstruction (2). The proportion of volvulus (8.3%) is higher than the rates reported in large Western datasets and closer to those described in regions where sigmoid volvulus is more prevalent (4,8). Regional differences in bowel habits, constipation burden, and anatomical predisposition have been proposed as explanations (4), although referral patterns to a tertiary center may also have contributed.

Of the 109 patients with volvulus, 4 (3.7%) had cecal volvulus. This sigmoid-dominant distribution differs from that of the US Nationwide Inpatient Sample and the classical Mayo Clinic series, in which cecal volvulus accounted for a substantially larger share of colonic volvulus (3,8,9). A low pretest probability of cecal volvulus in a sigmoid-dominant setting could influence how obstruction is initially interpreted; however, this hypothesis cannot be tested with the present data.

Current guidelines recommend endoscopic detorsion as the first-line approach for stable sigmoid volvulus, with resection reserved for failure, ischemia, or recurrence, whereas operative management is recommended for cecal volvulus (3,4,10). The management of our four patients was consistent with these recommendations: endoscopic detorsion was attempted in one patient before surgery but was unsuccessful; all four patients underwent resection. A single unsuccessful attempt cannot independently demonstrate the effectiveness or ineffectiveness of endoscopic treatment; the limited role of endoscopy in cecal volvulus is established by the existing literature rather than by

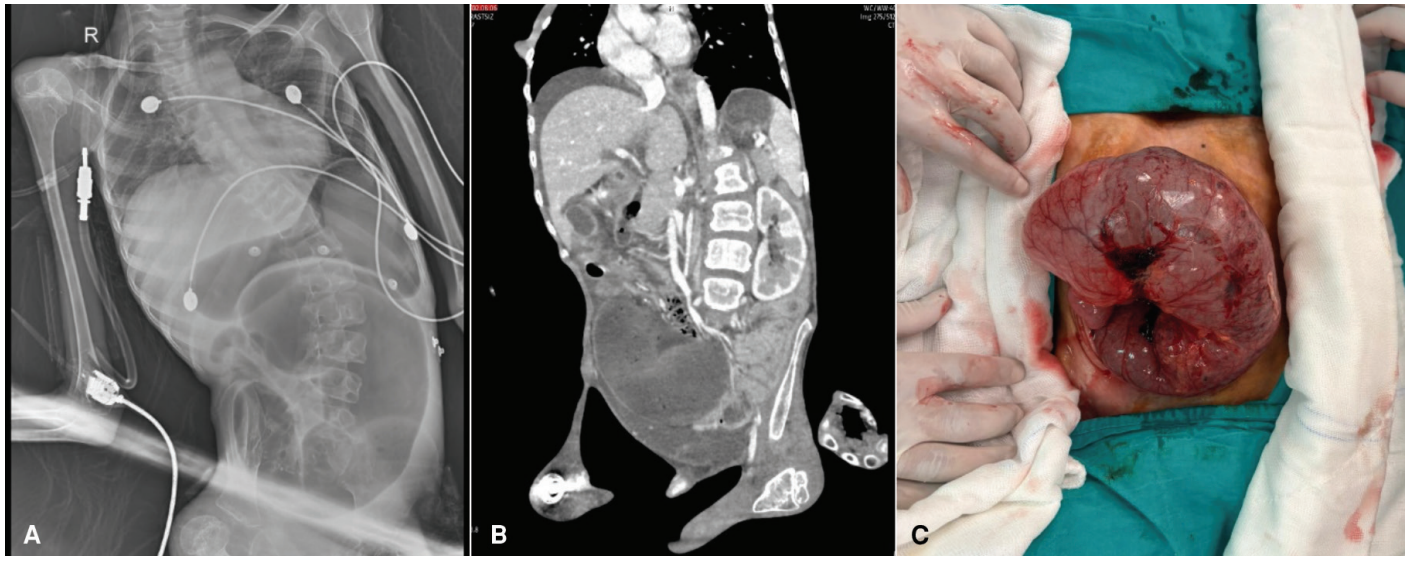


Figure 1. Representative radiological and intraoperative findings of cecal volvulus

(A) Plain abdominal radiograph showing a markedly dilated, ectopically positioned cecal loop suggestive of large-bowel obstruction. (B) Coronal abdominal computed tomography image demonstrating a displaced and distended cecum with findings compatible with cecal volvulus. (C) Intraoperative image showing a markedly distended and congested volvulized cecum, demonstrating the intraoperative appearance of cecal volvulus.

our data (3,10). The practical point illustrated by these cases is that, once volvulus is suspected, its anatomical site should be stated explicitly, because the recommended treatment pathways differ (4).

Three of the four patients had a neurological comorbidity (two had intellectual disability and one had epilepsy with cerebral palsy). Associations between cecal volvulus and neurological disease, impaired mobility, and altered bowel motility have been described in case reports (11-13). With four patients and no comparator group, our data cannot establish neurological comorbidity as a risk factor; this observation is reported solely to describe the case mix. The only patient without a neurological comorbidity was the oldest (83 years) and was also the only patient who died. Symptom duration before admission ranged from 1 to 4 days.

In two of the four patients, the original CT report did not specify cecal volvulus, describing obstructive or nonspecific ileus instead; in both patients, the diagnosis was established intraoperatively. The whirl sign was reported in two patients. Because the denominator was four and the secondary review was non-blinded, these observations do not constitute an assessment of diagnostic performance, sensitivity, or reliability of individual CT signs. Published series have described the CT features of cecal volvulus, including the whirl sign, ectopic cecal position, transition point, bird-beak configuration, and signs of ischemia (5,14,15), and have noted that obstruction may be reported without specification of the torsion site (6,16). In line with these reports and with recommendations on the structured communication of urgent imaging findings, explicitly stating the torsion site, the cecal position, and the presence or absence of ischemia in CT reports of suspected volvulus appears reasonable (17-19); the findings in our patients illustrate, but do not quantify, this point.

All four patients underwent open surgery. Laparoscopy was not selected because of the emergency setting, suspected ischemia, marked colonic distension, and limited working space (13). Laparoscopy may be feasible in selected stable patients, but open exploration is commonly preferred when bowel viability is uncertain (16).

Resection was performed in all four patients. A stoma without anastomosis was created in the two patients with ischemia and necrosis, and primary anastomosis was performed in the other two, reflecting individualized decisions based on bowel viability, physiology, and operative risk (20). Cecopexy has been associated with higher recurrence than resection in published series (21,22), and the choice among detorsion, cecopexy, cecostomy, and resection depends on bowel viability, patient condition, anatomy, and institutional practice (23). No recurrence was observed in our patients during follow-up.

One patient died within 90 days. This 83-year-old patient, classified as ASA 3E, was transferred from another hospital after a 3-day duration of symptoms, with an admission lactate of 4.04 mmol/L and a 15-cm gangrenous segment, and subsequently

died from postoperative cardiopulmonary complications after prolonged intubation. The most extensive gangrene in the series (70 cm) occurred in a 22-year-old patient who survived. A single death does not allow prognostic factors to be identified; larger series have reported advanced age, delayed presentation, bowel gangrene, and frailty as determinants of mortality after surgery for colonic volvulus (24-26).

We acknowledge that the principal clinical messages discussed above–the rarity of cecal volvulus, the importance of anatomical localization on CT, the different management of cecal and sigmoid volvulus, and the limited role of endoscopic detorsion in cecal volvulus–are already established in the literature and in current guidelines (3,4,10). The contribution of this study is therefore modest and descriptive. First, it provides a contemporary distribution of etiologies among 1.316 consecutive admissions for obstruction to a tertiary general surgery department in a region where sigmoid volvulus predominates. Second, it reports complete patient-level clinical, laboratory, radiological, operative, and outcome data for four cecal volvulus cases, together with their source denominator, which may be useful in future pooled analyses of this uncommon condition.

Study Limitations

This study has important limitations. First, only four patients had cecal volvulus; therefore, statistical inference was not possible, and all findings for this subgroup are descriptive and hypothesis-generating. Second, the study population comprised only patients who were admitted to the general surgery department at a single tertiary referral center; patients managed by other departments, discharged from the emergency department, or transferred elsewhere were not captured; therefore, the etiological frequencies may not apply to unselected emergency department populations. Third, the secondary CT review was non-blinded and conducted with knowledge of the operative diagnosis, introducing confirmation bias; the fact that two of four original reports did not specify cecal volvulus constitutes a case-level finding rather than a measure of diagnostic accuracy. Fourth, the 105 non-cecal volvulus cases were not subclassified by anatomical site, and detailed clinical and outcome data were not analyzed for these cases; therefore, no comparison between cecal and sigmoid volvulus could be made based on our data, and statements about their differing management rely on the existing literature. Fifth, because there was one death, two ischemic cases, and one endoscopic attempt, our interpretations regarding prognosis, CT reporting, and treatment selection are largely based on the literature. The retrospective single-center design limits generalizability. Prospective multicenter studies with structured CT reporting and standardized outcome capture are needed to characterize cecal volvulus more definitively.

Conclusion

Among 1.316 adults admitted to a general surgery department for intestinal obstruction, adhesive ileus was the most common etiology, and volvulus accounted for 8.3% of admissions. Cecal volvulus was identified in 4 of 109 patients with volvulus; all four underwent open resection, two required a stoma for ischemia and necrosis, and one 83-year-old patient died within 90 days.

The findings in the four patients with cecal volvulus are descriptive and hypothesis-generating, and do not permit conclusions regarding diagnostic performance, risk factors, or treatment effectiveness. They are consistent with existing guidelines that, once volvulus is suspected, its anatomical site should be explicitly determined and reported because the management of cecal and sigmoid volvulus differs. Larger prospective multicenter studies are needed.

Ethics

Ethics Committee Approval: The study was approved by the Clinical Research Ethics Committee of University of Health Sciences Türkiye, Ankara Etlik City Hospital (decision number: AEŞH-BADEK2-2025-745, date: 30.12.2025) and conducted in accordance with the Declaration of Helsinki.

Informed Consent: The requirement for informed consent was waived by the Clinical Research Ethics Committee due to the retrospective nature of the study.

Footnotes

Author Contributions

Surgical and Medical Practices: S.D., M.S.S., A.Y.K., F.Y., E.O., Concept: S.D., M.S.S., F.Y., B.B., L.S., Design: S.D., M.S.S., F.Y., H.D., Data Collection or Processing: S.D., M.S.S., A.Y.K., E.O., M.Ş., Analysis or Interpretation: S.D., A.Y.K., F.Y., B.B., M.Ş., L.S., Literature Search: S.D., M.S.S., A.Y.K., H.D., Writing: S.D., M.S.S., A.Y.K.

Conflict of Interest: Lukasz Szarpak is a member of the Editorial Board of the Eurasian Journal of Emergency Medicine. However, they were not involved in any stage of the editorial decision process for this manuscript. The editors who evaluated this manuscript are from different institutions. The other authors declared no conflict of interest.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Catena F, De Simone B, Coccolini F, Di Saverio S, Sartelli M, Ansaloni L. Bowel obstruction: a narrative review for all physicians. World J Emerg Surg. 2019;14:20.

2. Ten Broek RPG, Krielen P, Di Saverio S, Coccolini F, Biffl WL, Ansaloni L, et al. Bologna guidelines for diagnosis and management of adhesive small bowel obstruction (ASBO): 2017 update of the evidence-based guidelines from the world society of emergency surgery ASBO working group. World J Emerg Surg. 2018;13:24.
3. Alavi K, Poylin V, Davids JS, Patel SV, Felder S, Valente MA, et al. Prepared on behalf of the Clinical Practice Guidelines Committee of the American Society of Colon and Rectal Surgeons. The American Society of Colon and Rectal Surgeons clinical practice guidelines for the management of colonic volvulus and acute colonic pseudo-obstruction. Dis Colon Rectum. 2021;64:1046-57.
4. Tian BWCA, Vigutto G, Tan E, van Goor H, Bendinelli C, Abu-Zidan F, et al. WSES consensus guidelines on sigmoid volvulus management. World J Emerg Surg. 2023;18:34.
5. Delabrousse E, Sarlième P, Sailley N, Aubry S, Kastler BA. Cecal volvulus: CT findings and correlation with pathophysiology. Emerg Radiol. 2007;14:411-5.
6. Rosenblat JM, Rozenblit AM, Wolf EL, DuBrow RA, Den EI, Levsky JM. Findings of cecal volvulus at CT. Radiology. 2010;256:169-75.
7. Jackson P, Vigiola Cruz M. Intestinal obstruction: evaluation and management. Am Fam Physician. 2018;98:362-7.
8. Halabi WJ, Jafari MD, Kang CY, Nguyen VQ, Carmichael JC, Mills S, et al. Colonic volvulus in the United States: trends, outcomes, and predictors of mortality. Ann Surg. 2014;259:293-301.
9. Ballantyne GH, Brandner MD, Beart RW Jr, Ilstrup DM. Volvulus of the colon. Incidence and mortality. Ann Surg. 1985;202:83-92.
10. Naveed M, Jamil LH, Fujii-Lau LL, Al-Haddad M, Buxbaum JL, Fishman DS, et al. American Society for Gastrointestinal Endoscopy guideline on the role of endoscopy in the management of acute colonic pseudo-obstruction and colonic volvulus. Gastrointest Endosc. 2020;91:228-235. Erratum in: Gastrointest Endosc. 2020;91:721.
11. Lee EK, Kim JE, Lee YY, Kim S, Choi KH. A case of cecal volvulus presenting with chronic constipation in lissencephaly. Pediatr Gastroenterol Hepatol Nutr. 2013;16:131-4.
12. Takeuchi H, Ikeda Y, Komori Y, Tahara K, Shiromizu A, Hayashi H, et al. Cecal volvulus in cerebral palsy: report of a case. Surg Today. 2008;38:170-3.
13. Le CK, Nahirniak P, Qaja E. Cecal Volvulus. 2022 Sep 12. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022.
14. Dane B, Hindman N, Johnson E, Rosenkrantz AB. Utility of CT findings in the diagnosis of cecal volvulus. AJR Am J Roentgenol. 2017;209:762-6.
15. Macari M, Spieler B, Babb J, Pachter HL. Can the location of the CT whirl sign assist in differentiating sigmoid from caecal volvulus? Clin Radiol. 2011;66:112-7.
16. Madiba TE, Thomson SR. The management of cecal volvulus. Dis Colon Rectum. 2002;45:264-7.
17. Jacquemin Q, Coulier B, Rubay R. Acute volvulus of the cecum. J Belg Soc Radiol. 2020;104:40.
18. Nakamura Y, Kondo S, Narita K, Maeda S, Fonseca D, Honda Y, et al. Understanding CT imaging findings based on the underlying pathophysiology in patients with small bowel ischemia. Jpn J Radiol. 2023;41:353-66.
19. European Society of Radiology (ESR). ESR guidelines for the communication of urgent and unexpected findings. Insights Imaging. 2012;3:1-3.
20. O'Mara CS, Wilson TH Jr, Stonesifer GL, Cameron JL. Cecal volvulus: analysis of 50 patients with long-term follow-up. Ann Surg. 1979;189:724-31.
21. Gingold D, Murrell Z. Management of colonic volvulus. Clin Colon Rectal Surg. 2012;25:236-44.
22. Rabinovici R, Simansky DA, Kaplan O, Mavor E, Manny J. Cecal volvulus. Dis Colon Rectum. 1990;33:765-9.
23. Scott CD, Trotta BM, Dubose JJ, Ledesma E, Friel CM. A cruel twist: post-operative cecal volvulus. Ulus Travma Acil Cerrahi Derg. 2008;14:158-62.
24. Sylvester KR, Ooko PB, Mwachiro MM, Parker RK. Cecal volvulus in rural Kenya: delayed presentation contributes to high mortality. BMC Surg. 2021;21:430.
25. Easterday A, Aurit S, Driessen R, Person A, Krishnamurty DM. Perioperative outcomes and predictors of mortality after surgery for sigmoid volvulus. J Surg Res. 2020;245:119-26.
26. Ebrahimian S, Lee C, Tran Z, Sakowitz S, Bakhtiyar SS, Verma A, et al. Association of frailty with outcomes of resection for colonic volvulus: a national analysis. PLoS One. 2022;17:e0276917.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.63004
Eurasian J Emerg Med. 2026;25: 520-528

Blood Ethyl Alcohol Level, Clinical Outcomes and Number of Injured Body Regions in Young Adults after Traffic Accidents: A Retrospective Cohort Study

Abdussamed Vural[1], Turgut Dolanbay[2], Mustafa Oğuz Cumaoğlu[1], Sabri Sekme[1], Seyyid Rasim Yanmaz[1], Asaf Burak Vural[1]

[1]Niğde Ömer Halisdemir University Faculty of Medicine, Department of Emergency Medicine, Niğde, Türkiye
[2]Malatya Turgut Özal University Faculty of Medicine, Department of Emergency Medicine, Malatya, Türkiye

Abstract

Aim: To evaluate the association of blood ethyl alcohol concentration (BAC) with clinical outcomes, the number of injured anatomical regions and laboratory parameters in emergency department patients aged 20-40 years after traffic accidents.

Materials and Methods: This single-center retrospective cohort study included 451 patients whose BAC was measured between October 2022 and October 2023. BAC was categorized as 0, >0-50, or >50 mg/dL (legal limit). Group comparisons were assessed using the chi-square test, Fisher's exact test, the Mann-Whitney U test; correlations were assessed using Spearman's rho. Extremity involvement was analyzed with logistic regression adjusted for age and sex. All analyses were exploratory, with false discovery rate correction applied.

Results: Measurable alcohol was detected in 131 patients (29.0%), and 79 patients (17.5%) were above the legal limit. Males predominated in the above-limit group (93.7% vs. 78.0%; p=0.002). No statistically significant associations were observed for mortality [1.3% vs. 2.2%; odds ratio (OR) 0.58, 95% confidence interval (CI) 0.07-4.73], hospital admission (OR 1.59, 95% CI 0.93-2.73), surgical intervention (OR 1.31, 95% CI 0.64-2.67), or extremity involvement (adjusted OR 1.71, 95% CI 0.99-2.97). BAC was correlated weakly with the number of injured regions (rho=0.204, 95% CI 0.114-0.291), alanine aminotransferase (rho=0.297), and aspartate aminotransferase (rho=0.341) (all q<0.001). The number of injured regions increased across alcohol categories (Jonckheere-Terpstra p<0.001), driven mainly by differences between patients with and without measurable alcohol.

Conclusion: In young adults following traffic accidents, BAC was not significantly associated with mortality, admission, surgery, or extremity involvement, although the statistical power to detect an association with mortality was limited. Measurable alcohol was weakly associated with a greater number of injured regions and higher liver enzyme levels. These exploratory findings require prospective confirmation with validated injury-severity scores.

Keywords: Clinical outcome, emergency department, ethyl alcohol, traffic accidents, wounds and injuries

Introduction

Ethyl alcohol consumption remains a serious public health concern, given its significant impact on traffic accidents, which are among the leading causes of death and disability worldwide (1,2). Research in this field aims to clarify the relationship between ethanol consumption and traffic accidents. Alcohol consumption is known to reduce attention, prolong reaction times in response to unexpected situations, and delay motor responses. It is also thought to increase the risk of traffic accidents by promoting behaviors such as non-compliance with traffic regulations (3). Previous studies support this relationship, showing that individuals who consume alcohol are at higher risk of being involved in accidents (4-6).

The frequency of alcohol-related traffic accidents varies across age

Corresponding Author: Abdussamed Vural, MD, Department of Emergency Medicine, Niğde Training and Research Hospital, Niğde, Turkey
E-mail: abdussamedvural@gmail.com **ORCID ID:** orcid.org/0000-0003-4506-916X

Cite this article as: Vural A, Dolanbay T, Cumaoğlu MO, Sekme S, Yanmaz SR, Vural AB. Blood ethyl alcohol level, clinical outcomes and number of injured body regions in young adults after traffic accidents: a retrospective cohort study Eurasian J Emerg Med. 2026;25: 520-528

Received: 29.04.2026
Accepted: 31.10.2026
Published: 09.11.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

groups, with younger individuals being more susceptible to such accidents, and younger age being more strongly associated with fatal traffic accidents (4,7,8). According to Erik Erikson's theory of psychosocial development, individuals aged 20-40 years are in young adulthood, the first period of adulthood. During this period, individuals may engage in risky behaviors while trying to overcome the challenges of the transition from adolescence to adulthood (9,10). Despite the existing body of knowledge, studies evaluating the relationship between ethyl alcohol use and traffic accidents specifically in young adulthood (20-40 years) - at the level of clinical outcomes, the distribution of injuries and laboratory parameters - remain limited, and recent registry-based studies have reported inconsistent associations between alcohol, injury severity and mortality (11,12).

The aim of this study was to evaluate the association between blood ethyl alcohol level and mortality, hospital admission, need for surgical intervention, length of hospital stay, number of injured anatomical regions, and laboratory parameters in patients aged 20-40 years presenting to the emergency department after traffic accidents. We hypothesized that blood ethyl alcohol level would not be associated with mortality in this population but might be associated with the number of injured regions and with liver function tests.

MATERIALS AND METHODS

Study Design and Setting

This single-center retrospective cohort study reviewed the records of patients presenting to the Emergency Medicine Department of Niğde Ömer Halisdemir University after traffic accidents. Data on patients aged 20-40 years presenting with a diagnostic code of traffic accident (ICD-10: V01-V99) between 01.10.2022 and 01.10.2023 were retrieved from the Hospital Information Management System (Karmed). All eligible patients during the study period were included; no a priori sample size calculation was performed, and, therefore, the precision of the main estimates is reported with 95% confidence intervals. The study protocol was not registered in advance, and all analyses are considered exploratory.

This study was approved by the Niğde Ömer Halisdemir University Non-interventional Clinical Research Ethics Committee (decision date: 28.12.2023, protocol no: 2023/113). As this study was a retrospective chart review, informed consent was not obtained from patients.

Inclusion and Exclusion Criteria

All patients aged 20-40 years who presented to the emergency department after a traffic accident during the specified date range and had blood ethyl alcohol levels measured were included in the study. The following were excluded: (1) patients who refused examination or work-up or who left the emergency department against medical advice (n=3); (2) a record with no clinical or laboratory data (n=1); (3) a patient who presented solely for forensic blood alcohol testing and was not a traffic-accident victim (n=1); and (4) patients whose blood ethyl alcohol level was not measured (n=8). Of the 464 patients initially identified, 13 were excluded on these grounds, resulting in a final analysis cohort of 451 patients (Figure 1).

Variables

Sex, age, blood ethyl alcohol level, hospital admission, length of hospital stay, need for surgical intervention, in-hospital mortality, involvement of six anatomical regions (head/brain, cervical spine, thorax, abdomen, pelvis, and extremities), and baseline laboratory parameters [hemoglobin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), urea, creatinine] were recorded. Ethyl alcohol concentration was measured quantitatively in blood samples obtained in the emergency department; a value of 0 mg/dL indicated no detectable ethanol. Patients were classified into three mutually exclusive categories: no measurable alcohol (0 mg/dL); below the legal limit (>0-50 mg/dL); and above the legal limit (>50 mg/dL; 0.5‰). For two-group comparisons, patients above the legal limit were compared with all patients at or below the limit (≤50 mg/dL, including those with no measurable alcohol).

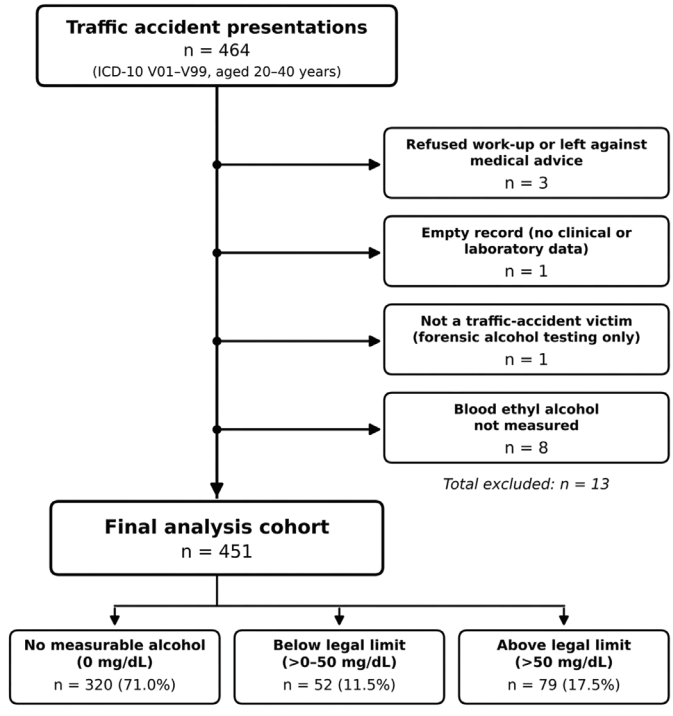


Figure 1. Patient flow diagram

Ascertainment of Injured Regions

Involvement of each anatomical region was determined from the documented clinical and/or radiological evaluations of the region in the medical record. Evaluation was at the discretion of the treating physician and was not protocolized. Regions without a documented evaluation were coded as not involved. This coding was used for both region-specific comparisons and the total number of injured regions, so all region-level analyses include the full cohort (n=451). No regional evaluation was documented for 78 patients (17.3%). The number of injured regions (range 0-6) is a descriptive count of affected body regions and not equivalent to validated injury-severity measures such as the Injury Severity Score (ISS) or the New Injury Severity Score (NISS). As a sensitivity analysis, region-specific comparisons were repeated among patients with a documented evaluation of the corresponding region.

Statistical Analysis

Normality of distribution was assessed using the Shapiro-Wilk (n≤50) or Kolmogorov-Smirnov (n>50) tests. Non-normally distributed continuous variables were presented as median (minimum-maximum) and categorical variables as number (n) and percentage (%). For two-group comparisons, the Mann-Whitney U test was used for continuous variables and the chi-square test with Yates' continuity correction was used for categorical variables; Fisher's exact test was used instead for 2x2 tables in which any expected cell count was below 5. Odds ratios (ORs) with 95% confidence intervals (CIs) are reported for binary outcomes. Spearman's correlation analysis was used to assess associations between blood ethyl alcohol level (analyzed as a continuous variable) and other continuous variables; 95% CIs were obtained with the Fisher z transformation, and partial Spearman's correlations adjusted for the number of injured regions were calculated for ALT and AST. The potential ordered relationship between the three-category blood alcohol level and the number of injured regions was evaluated using the Jonckheere-Terpstra trend test, followed by pairwise Mann-Whitney U tests. The association between alcohol status (>50 vs. ≤50 mg/dL) and extremity involvement was examined with multivariable logistic regression adjusted for age and sex, the only pre-injury covariates available in the dataset; injury-related variables (other injured regions, surgery, length of stay, laboratory values) were not included because they are consequences of the injury rather than confounders. Multicollinearity, linearity of the logit for age, and model fit were assessed with variance inflation factors (threshold >5), the Box-Tidwell test, and the Hosmer-Lemeshow test and Nagelkerke R^2, respectively. With 100 events and three parameters, there were approximately 33 events per variable. Exploratory sensitivity analyses modeled age with a cubic B-spline and alcohol either in three categories or as a continuous variable, and restricted the analysis to patients with a documented extremity evaluation. Because no protocol was registered, all analyses were considered exploratory. To account for multiple testing, Benjamini-Hochberg false-discovery-rate q values were calculated across all 19 inferential tests, with Holm-adjusted p values used as a more conservative sensitivity analysis (Supplementary Table S1). Missing laboratory data (16.9-18.0%) were not imputed, and analyses involving laboratory variables used all available observations. Analyses were performed using SPSS 27.0 (IBM Corp., Armonk, NY, USA) and R version 4.3.3 (R Foundation for Statistical Computing, Vienna, Austria). A two-sided p value <0.05 was considered statistically significant.

Results

Of the 464 patients who presented to the emergency department following traffic accidents during the study period, 13 were excluded, and 451 were included in the final analysis (Figure 1). The descriptive characteristics of the final analysis cohort are summarized in Table 1. Of the 451 patients, 364 (80.7%) were male and 87 (19.3%) were female; the median age was 27 years (range, 20-40). Regarding blood ethyl alcohol level, 320 patients (71.0%) had no measurable blood ethyl alcohol; 52 patients (11.5%) had a level below the legal limit (>0-50 mg/dL); and 79 patients (17.5%) had a level above the legal limit (>50 mg/dL). Accordingly, measurable blood ethyl alcohol was detected in 131 patients (29.0%). Among the 131 alcohol-positive patients, the median blood ethyl alcohol level was 106.7 mg/dL (range, 0.01-327.2); in the above-legal-limit group, the median blood ethyl alcohol level was 164.9 mg/dL (range, 54.6-327.2). Overall, 9 patients (2.0%) died, 104 (23.1%) were admitted to the hospital, and 52 (11.5%) underwent surgical intervention. The median length of hospital stay was 0 days (range, 0-62), reflecting that most patients were discharged on the same day (Table 1). Among the 52 patients who underwent surgical intervention, orthopedic surgery was performed most frequently (n=41, 78.8%), followed by neurosurgery (n=5, 9.6%); the remaining patients underwent plastic surgery (n=3), general surgery (n=1), or a combined general surgery-orthopedics procedure (n=1), and the surgical specialty was not recorded for one patient.

Comparisons between patients with a blood ethyl alcohol level above the legal limit and those at or below the limit are presented in Table 2. The proportion of males was higher in the above-legal-limit group (93.7% vs. 78.0%; OR 4.18, 95% CI 1.64-10.69; p=0.002), whereas median age did not differ between groups [28 (20-40) vs. 27 (20-40) years; p=0.758]. The number of injured regions was higher in the above-legal-limit group (median 1, range 0-4) than in the comparison group (median

Table 1. Descriptive characteristics of the final analysis cohort (n=451)	
Variable	**Value**
Total number of patients	451
Sex, male/female, n (%)	364 (80.7%)/87 (19.3%)
Age, years, median (min-max)	27 (20-40)
Blood ethyl alcohol level, n (%)	
No measurable alcohol (0 mg/dL)	320 (71.0%)
Below legal limit (>0-50 mg/dL)	52 (11.5%)
Above legal limit (>50 mg/dL)	79 (17.5%)
Blood ethyl alcohol level among alcohol-positive patients (n=131), mg/dL, median (min-max)	106.7 (0.01-327.2)
Number of injured regions, median (min-max)	1 (0-4)
Hospital admission, n (%)	104 (23.1%)
Surgical intervention, n (%)	52 (11.5%)
Mortality, n (%)	9 (2.0%)
Length of hospital stay, days, median (min-max)	0 (0-62)
Hemoglobin, g/dL, median (min-max) (n=375)	15.4 (7.3-19.5)
ALT, U/L, median (min-max) (n=370)	23.5 (2-861)
AST, U/L, median (min-max) (n=370)	25 (11-642)
Urea, mg/dL, median (min-max) (n=372)	27 (8-68)
Creatinine, mg/dL, median (min-max) (n=372)	0.85 (0.40-6.00)

Percentages were calculated based on the total cohort (n=451); laboratory variables are based on available data (n shown). Blood ethyl alcohol level categories are mutually exclusive (totals sum to 100.0%).
ALT: Alanine aminotransferase, AST: Aspartate aminotransferase, min-max: Minimum-maximum

Table 2. Comparisons according to the legal blood ethyl alcohol limit (>50 vs. ≤50 mg/dL)					
Variable	**Above legal limit (n=79)**	**≤50 mg/dL[a] (n=372)**	**OR (95% CI)**	**p**	**q**
Male sex, n (%)	74 (93.7%)	290 (78.0%)	4.18 (1.64-10.69)	0.002	-
Age, years, median (min-max)	28 (20-40)	27 (20-40)	-	0.758	-
Number of injured regions, median (min-max)	1 (0-4)	0 (0-4)	-	0.002	0.004
Hospital admission, n (%)	24 (30.4%)	80 (21.5%)	1.59 (0.93-2.73)	0.120	0.190
Length of stay among admitted patients, days, median (min-max)[b]	4 (1-62)	5 (1-28)	-	0.591	0.721
Mortality, n (%)	1 (1.3%)	8 (2.2%)	0.58 (0.07-4.73)	1.000[c]	1.000
Surgical intervention, n (%)	11 (13.9%)	41 (11.0%)	1.31 (0.64-2.67)	0.589	0.721

The chi-square test with Yates' continuity correction was used for categorical variables and the Mann-Whitney U test for continuous variables. q: Benjamini-Hochberg false-discovery-rate adjusted p value across 19 tests; -: not part of the adjusted test family. [a]The ≤50 mg/dL group includes patients with no measurable alcohol (0 mg/dL, n=320) and those with a level below the legal limit (>0-50 mg/dL, n=52). [b]Only patients who were admitted (above legal limit n=24; ≤50 mg/dL n=80); the upper bound of 62 days corresponds to a single case requiring multiple trauma management and orthopedic surgery. [c]Fisher's exact test.
OR: Odds ratio, CI: Confidence interval

0, range 0-4) (p=0.002; q=0.004). No statistically significant association was observed between a blood ethyl alcohol level above the legal limit and hospital admission (30.4% vs. 21.5%; OR 1.59, 95% CI 0.93-2.73; p=0.120), mortality (1.3% vs. 2.2%; OR 0.58, 95% CI 0.07-4.73; Fisher's exact test, p=1.000), or need for surgical intervention (13.9% vs. 11.0%; OR 1.31, 95% CI 0.64-2.67; p=0.589). Among admitted patients, no statistically significant difference in length of stay was observed between the two groups (median 4 vs. 5 days; p=0.591), and the upper bound of 62 days in the above-legal-limit group corresponded to a single case requiring management of multiple trauma and orthopedic surgery.

The Spearman's correlation analyses of blood ethyl alcohol level (evaluated as a continuous variable) with other continuous variables are summarized in Table 3 and Supplementary Figure S1. Blood ethyl alcohol level showed a weak positive correlation

with the number of injured regions (n=451; rho=0.204, 95% CI 0.114-0.291; q<0.001), and weak-to-moderate positive correlations with ALT (n=370; rho=0.297, 95% CI 0.201-0.387; q<0.001) and AST (n=370; rho=0.341, 95% CI 0.248-0.428; q<0.001). After adjustment for the number of injured regions, the partial correlation coefficients were 0.259 for ALT and 0.289 for AST (both p<0.001). Blood ethyl alcohol level also correlated with hemoglobin (rho=0.195; q<0.001), urea (rho=-0.167; q=0.004), and length of hospital stay (rho=0.126; q=0.017), but not with creatinine (rho=0.081; p=0.117) or age (rho=0.046; p=0.330) (Table 3).

In terms of injured regions, the most frequently affected in the overall cohort were the extremities (n=100, 22.2%) and the head/brain (n=94, 20.8%), followed by the thorax (n=47, 10.4%), abdomen (n=21, 4.7%), pelvis (n=16, 3.5%), and cervical region (n=10, 2.2%). Region-specific comparisons between the above-legal-limit group and the comparison group are presented in Table 4. Head or brain involvement was more frequent in the above-legal-limit group (30.4% vs. 18.8%; OR 1.88, 95% CI 1.09-3.25; p=0.032), but this difference was not statistically significant after false discovery rate correction (q=0.067). Extremity involvement was 30.4% vs. 20.4% (OR 1.70, 95% CI 0.99-2.92; p=0.074; q=0.141). No statistically significant differences were observed for the cervical, thoracic, abdominal or pelvic regions (all p≥0.499). Documented evaluation of the head/brain region was more frequent in the above-legal-limit group (39.2% vs. 28.2%; p=0.059), whereas evaluation rates for the other regions were similar between groups (Supplementary Table S2). Among patients with a documented head/brain evaluation, involvement did not differ between groups (77.4% vs. 66.7%; p=0.359).

In multivariable logistic regression of the full cohort (n=451; 100 events), a blood ethyl alcohol level above the legal limit was not significantly associated with extremity involvement after adjustment for sex and age (adjusted OR 1.71, 95% CI 0.99-2.97; p=0.056) (Table 5). There was no evidence of multicollinearity (variance inflation factors; 1.00-1.03). Model calibration was adequate (Hosmer-Lemeshow p=0.742), but explained variance was low (Nagelkerke R^2=0.020). The Box-Tidwell test indicated non-linearity of the logit for age (p=0.008). When age was modeled as a cubic B-spline, the estimate was similar (adjusted OR 1.65, 95% CI 0.94-2.88). In exploratory sensitivity analyses (Supplementary Table S3), the association was statistically significant under three conditions: when the analysis was restricted to the 213 patients with a documented extremity evaluation (adjusted OR 2.86, 95%

Table 3. Spearman's correlations between blood ethyl alcohol level and continuous variables

Variable	n	rho (95% CI)	p	q
Number of injured regions	451	0.204 (0.114 to 0.291)	<0.001	<0.001
ALT	370	0.297 (0.201 to 0.387)	<0.001	<0.001
AST	370	0.341 (0.248 to 0.428)	<0.001	<0.001
Hemoglobin	375	0.195 (0.096 to 0.291)	<0.001	<0.001
Urea	372	-0.167 (-0.264 to -0.066)	0.001	0.004
Creatinine	372	0.081 (-0.020 to 0.182)	0.117	0.190
Length of hospital stay	451	0.126 (0.034 to 0.216)	0.007	0.017
Age	451	0.046 (-0.047 to 0.138)	0.330	-

All correlations were calculated using Spearman's rho; 95% CIs were obtained with the Fisher z transformation. Pairwise complete observations were used. q: Benjamini-Hochberg false-discovery-rate adjusted p value across 19 tests. Partial correlations adjusted for the number of injured regions: ALT rho=0.259, AST rho=0.289 (both p<0.001). The full correlation matrix is shown in Supplementary Figure S1.
CI: Confidence interval, ALT: Alanine aminotransferase, AST: Aspartate aminotransferase

Table 4. Involvement of anatomical regions according to the legal blood ethyl alcohol limit (n=451)

Region	Above legal limit (n=79), n (%)	≤50 mg/dL (n=372), n (%)	OR (95% CI)	p	q
Head/brain	24 (30.4%)	70 (18.8%)	1.88 (1.09-3.25)	0.032	0.067
Extremities	24 (30.4%)	76 (20.4%)	1.70 (0.99-2.92)	0.074	0.141
Thorax	10 (12.7%)	37 (9.9%)	1.31 (0.62-2.76)	0.607	0.721
Pelvis	4 (5.1%)	12 (3.2%)	1.60 (0.50-5.10)	0.499[a]	0.721
Abdomen	4 (5.1%)	17 (4.6%)	1.11 (0.36-3.40)	0.773[a]	0.864
Cervical	1 (1.3%)	9 (2.4%)	0.52 (0.06-4.14)	1.000[a]	1.000

All 451 patients are included; regions without a documented clinical or radiological evaluation were coded as not involved. The chi-square test with Yates' continuity correction was used when expected cell counts were sufficient; [a]Fisher's exact test was used when any expected cell count was below 5. q: Benjamini-Hochberg false-discovery-rate adjusted p value across 19 tests.
CI: Confidence interval, OR: Odds ratio

CI 1.31-6.24); when alcohol was modeled in three categories (>0-50 mg/dL: adjusted OR 2.50, 95% CI 1.31-4.77; >50 mg/dL: adjusted OR 2.00, 95% CI 1.14-3.53; reference 0 mg/dL); and when blood ethyl alcohol level was modeled as a continuous variable (adjusted OR 1.04 per 10 mg/dL, 95% CI 1.01-1.07). When blood alcohol levels were categorized into three groups (none/below legal limit/above legal limit), an ordered increase in the number of injured regions was observed (Jonckheere-Terpstra trend test, z=3.81; p<0.001; Figure 2). Pairwise comparisons showed differences between the no-alcohol group and both the >0-50 mg/dL group (p=0.004) and the >50 mg/dL group (p<0.001), but not between the >0-50 mg/dL and >50 mg/dL groups (p=0.994).

Discussion

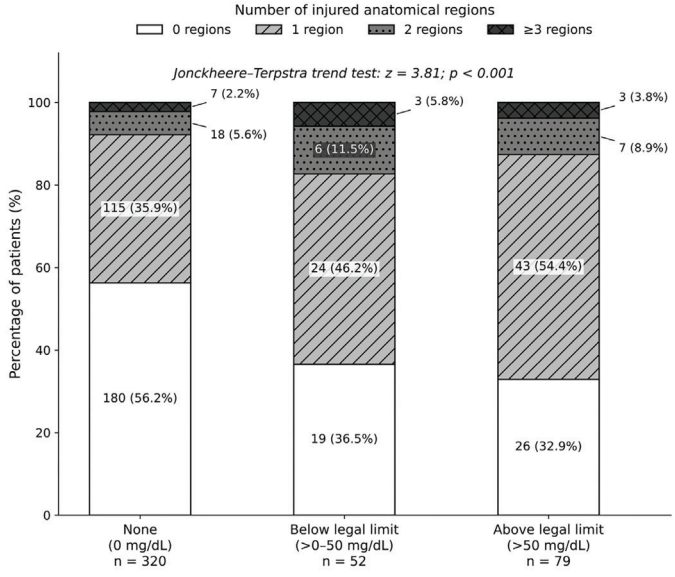


Figure 2. Distribution of the number of injured anatomical regions by blood ethyl alcohol category. Bars show the number (percentage) of patients with 0, 1, 2, or ≥3 injured anatomical regions within each blood ethyl alcohol category. The Jonckheere-Terpstra trend test, a rank-based nonparametric test, was used given the non-normal, discrete distribution of this variable. The distribution shift occurs mainly between the absence and presence of measurable alcohol, with no statistically significant difference between the >0-50 and >50 mg/dL groups

In this study of 451 patients aged 20-40 years presenting to the emergency department after traffic accidents, measurable blood ethyl alcohol was detected in 29.0% of patients, and a level above the legal limit (50 mg/dL) was detected in 17.5%. Blood ethyl alcohol level was weakly associated with the number of injured anatomical regions and with ALT and AST levels; whereas no statistically significant association was observed with mortality, hospital admission, surgical intervention, or involvement of any specific body region after correction for multiple testing. The alcohol-positive rates fall within the wide range of 8% to 47.1% reported in international trauma centers (13,14); this variability may stem from differences in screening policies, patient populations, and cultural/legal factors. A study conducted in Turkey reported this rate as 22.1% (15).

The higher proportion of males in the above-legal-limit group (93.7% vs. 78.0%) is consistent with behavioral studies showing that driving under the influence of alcohol is markedly more common among young male drivers; one study reported that 18.5% of young male drivers and 11.8% of young female drivers had driven after drinking alcohol (16). In contrast, the absence of an age difference between the two groups (p=0.758), when considered alongside epidemiological data showing that the risk of late-night accidents rises markedly among individuals aged 18-24 (17), suggests that restricting our sample from the outset to a narrow age range (20-40 years) may have reduced age-related variance.

The absence of a statistically significant association between blood ethyl alcohol level and mortality should be interpreted with caution because only nine deaths occurred, and the confidence interval was wide (OR 0.58, 95% CI 0.07-4.73), which is compatible with both a substantial reduction and a more than fourfold increase in the odds of death. Previous studies have reported heterogeneous findings. Jurkovich et al. (13) reported that alcohol intoxication altered the initial clinical assessment of 2237 trauma patients but did not directly predict outcome. Hsieh et al. (18) showed that alcohol intoxication did not increase in-hospital mortality and was associated with a lower ISS in some subgroups. Similarly, an analysis of 88,794 patients from the National Trauma Data Bank found no significant difference in

Table 5. Logistic regression analysis for extremity involvement (n=451; 100 events)

Variable	Crude OR (95% CI)	p	Adjusted OR (95% CI)	p
Blood ethyl alcohol >50 mg/dL (ref. ≤50 mg/dL)	1.70 (0.99-2.92)	0.055	1.71 (0.99-2.97)	0.056
Age, per year	0.97 (0.93-1.01)	0.132	0.97 (0.93-1.01)	0.126
Male sex (ref. female)	1.11 (0.63-1.98)	0.711	1.00 (0.56-1.80)	0.993

Binary logistic regression. Regions without a documented evaluation were coded as not involved. Model diagnostics: variance inflation factors 1.00-1.03; events per variable ≈33; Hosmer-Lemeshow χ^2=5.14, df=8, p=0.742; Nagelkerke R^2=0.020; area under the ROC curve 0.57. The Box-Tidwell test indicated non-linearity of the logit for age (p=0.008); with age modeled as a cubic B-spline (3 df), the adjusted OR for blood ethyl alcohol >50 mg/dL was 1.65 (95% CI 0.94-2.88; p=0.078).
OR: Odds ratio, CI: Confidence interval (wald)

in-hospital mortality according to alcohol positivity (19), and an analysis of 6,268 patients from the TraumaRegister DGU® likewise found no difference in overall mortality between groups (20). Other studies, including a recent analysis of the Japan Trauma Data Bank, have reported lower mortality among intoxicated or drinking patients (12,21), a finding generally attributed to lower injury severity in these patients or to residual confounding, and a recent single-center registry study reported inconsistent associations between blood alcohol and mortality (11). Because injury severity was not measured in our study, we could not assess whether these mechanisms applied to our cohort.

No statistically significant difference was observed between groups in the need for surgical intervention (OR 1.31, 95% CI 0.64-2.67), consistent with the mixed picture in the literature: some studies have shown alcohol positivity to be associated with increased resource use (ICU admission, consultation, procedural intervention) (22,23), while others have found no significant difference (14,24). Similarly, no statistically significant difference was observed in hospital admission (30.4% vs. 21.5%; OR 1.59, 95% CI 0.93-2.73) or in length of stay among patients who were admitted (p=0.591), consistent with Beydoun et al.'s (14) finding that alcohol positivity did not directly predict length of hospital stay and with another study that found no significant association between AUDIT/BAC and hospital/ICU length of stay (25). The confidence intervals for these outcomes do not exclude modest associations.

In our study, blood ethyl alcohol level showed a weak positive correlation with the number of injured anatomical regions (rho=0.204). In the full cohort, extremity involvement was not significantly associated with blood ethyl alcohol levels above the legal limit after adjustment for sex and age (adjusted OR 1.71, 95% CI 0.99-2.97). Larger estimates were obtained in the analysis restricted to patients with a documented extremity evaluation and in the categorical analysis, but these sensitivity analyses were exploratory, and the restricted analysis was susceptible to ascertainment bias. Beydoun et al. (14), using a propensity-score approach, reported that higher blood alcohol level was positively associated with head injury but inversely associated with extremity injuries; this could not be examined further because accident mechanism was not recorded in our dataset. Accident mechanism, road-user role, use of protective devices, and timing of blood sampling were not available, so residual confounding cannot be excluded in either direction.

Blood ethyl alcohol level showed weak-to-moderate positive correlations with ALT (rho=0.297, 95% CI 0.201-0.387) and AST (rho=0.341, 95% CI 0.248-0.428); these correlations persisted after adjustment for the number of injured regions. Kelley et al. (26) reported that blood alcohol concentration was associated with differences in hematologic and serum chemistry parameters in trauma patients. However, AST is also released from injured skeletal muscle; the number of injured regions does not measure soft-tissue damage; and the timing of blood sampling, chronic alcohol use, and pre-existing liver disease were not recorded. These correlations therefore cannot be attributed to a direct hepatic effect of acute alcohol exposure. The positive correlation with hemoglobin and the negative correlation with urea were not hypothesized, and should be regarded as exploratory.

The Jonckheere-Terpstra test indicated an ordered shift in the number of injured regions across the three alcohol categories (z=3.81; p<0.001). However, the distributions of the >0-50 and >50 mg/dL groups did not differ (p=0.994), and the principal difference lay between patients with and without measurable alcohol levels (Figure 2). This exploratory pattern should not be interpreted as evidence of a dose-response or threshold relationship.

Head/brain involvement was more frequent in the above-legal-limit group of the full cohort (30.4% vs. 18.8%); however, the difference was not statistically significant after correction for multiple testing and was absent among patients with a documented head/brain evaluation (77.4% vs. 66.7%; p=0.359). Because a head or brain evaluation was documented more often in patients above the legal limit (39.2% vs. 28.2%), this difference likely reflects differential evaluation rather than differences in injury occurrence; impaired consciousness in intoxicated patients may lower the threshold for head imaging. Rundhaug et al. (24) showed that blood alcohol level lowered the Glasgow Coma Scale score in patients with moderate-to-severe traumatic brain injury but was not associated with computed tomography findings, and Leijdesdorff et al. (27) reported less frequent cerebral injury and lower injury severity among intoxicated patients with traumatic brain injury. In a study of 347 severely injured trauma patients, Riuttanen et al. (28) reported that alcohol positivity did not significantly affect ISS or ICU length of stay. Finally, studies showing that very high blood alcohol levels (>200-300 mg/dL) are associated with increased resource use (23,29) could not be examined in our cohort, because only 19 patients had levels above 200 mg/dL.

Study Limitations

The main limitations of our study are as follows: (1) the single-center, retrospective design, which may have led to data loss and selection bias; and the use of a convenience sample without an a priori sample size calculation. (2) The low number of mortality events (n=9) provides very limited power for this outcome; therefore, the absence of a statistically significant association should not be interpreted as evidence of no association. (3)

the absence of laboratory data in a subset of patients (16.9-18.0%), with ALT/AST missing more often in the comparison group (19.1%) than in the above-legal-limit group (12.7%); (4) The lack of a validated ISS (such as ISS or NISS): the number of injured regions is a descriptive count that does not capture injury severity and limits direct comparison with ISS/NISS-based studies such as those by Beydoun et al. (14) and Riuttanen et al. (28). (5) Regional involvement was ascertained from clinically indicated evaluations; regions without a documented evaluation were coded as not involved, which may have led to misclassification of some injuries; head/brain evaluation was documented more often in patients above the legal limit, introducing possible ascertainment bias; (6) accident mechanism, road-user role (vehicle occupant/pedestrian/motorcyclist), protective-device use, and time from injury to blood sampling were not recorded; therefore, the regression model was adjusted only for sex and age, and residual confounding is likely (7). Blood ethyl alcohol testing in the emergency department is applied selectively in clinical practice; it is not routinely ordered for pedestrian trauma but is more frequently requested for patients who were driving a vehicle, which suggests that our sample may carry a selection bias favoring driver-related cases that could not be quantified. (8) the inability to distinguish whether the alcohol-liver enzyme association reflects acute or chronic use, and the possibility that AST partly reflects skeletal muscle injury; (9) the restriction of the sample to the 20-40-year age range, which limits generalizability to broader age groups; and (10) The absence of a registered protocol meant that all analyses were exploratory; although multiplicity was addressed with false-discovery-rate correction, the findings should be regarded as hypothesis-generating.

Conclusion

In young adults presenting to the emergency department after traffic accidents, no statistically significant association was observed between blood ethyl alcohol level and mortality, hospital admission, surgical intervention, or involvement of any specific body region, although the study had limited power to detect differences in mortality. Measurable blood ethyl alcohol was weakly associated with a greater number of injured anatomical regions and higher liver enzyme levels. These exploratory findings require confirmation in prospective, multicenter studies using validated ISS and complete data on accident mechanisms.

Ethics

Ethics Committee Approval: This study was approved by the Niğde Ömer Halisdemir University Non-interventional Clinical Research Ethics Committee (decision date: 28.12.2023, protocol no: 2023/113).

Informed Consent: As this study was a retrospective chart review, informed consent was not obtained from patients.

Authorship Contributions

Concept: A.V., T.D., Design: A.V., M.O.C., Data Collection or Processing: A.V., M.O.C., S.S., S.R.Y., Analysis or Interpretation: A.V., T.D., A.B.V., Literature Search: S.S., S.R.Y., A.B.V., Writing: A.V., T.D., M.O.C., S.R.Y.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

Supplementary Tables Link: https://d2v96fxpocvxx.cloudfront.net/bb2eeae3-0e60-42a4-acea-81e4a349912c/content-images/a7b94b87-94e5-415e-ba1a-90fa625aaa34.pdf

Supplementary Figures Link: https://d2v96fxpocvxx.cloudfront.net/bb2eeae3-0e60-42a4-acea-81e4a349912c/content-images/fd635668-366b-4f72-9e12-2d336e53a6ab.pdf

References

1. Reynaud M, Le Breton P, Gilot B, Vervialle F, Falissard B. Alcohol is the main factor in excess traffic accident fatalities in France. Alcohol Clin Exp Res. 2002;26:1833-9.
2. Rönz K, Hirschi T, Becker S, Krummrey G, Exadaktylos AK, Sauter T, et al. Predictors of high resource consumption in alcohol intoxicated patients in the emergency department. Int J Environ Res Public Health. 2020;17:4122.
3. Nishitani Y. Alcohol and traffic accidents in Japan. IATSS Res. 2019;43:79-83.
4. Papalimperi AH, Athanaselis SA, Mina AD, Papoutsis II, Spiliopoulou CA, Papadodima SA. Incidence of fatalities of road traffic accidents associated with alcohol consumption and the use of psychoactive drugs: a 7-year survey (2011-2017). Exp Ther Med. 2019;18:2299-306.
5. Taylor B, Rehm J. The relationship between alcohol consumption and fatal motor vehicle injury: high risk at low alcohol levels. Alcohol Clin Exp Res. 2012;36:1827-34.
6. Porto DE, Cavalcanti YW, Forte FDS. Maxillofacial trauma due to traffic accidents and falls: an exploratory study of associated factors. Med Oral Patol Oral Cir Bucal. 2021;26:e349-56.
7. Qirjako G, Burazeri G, Hysa B, Roshi E. Factors associated with fatal traffic accidents in Tirana, Albania: cross-sectional study. Croat Med J. 2008;49:734-40.
8. Zewde T. Determinants that lead drivers into traffic accidents: a case of Arba Minch City, South Ethiopia. Sci J Appl Math Stat. 2017;5:210.
9. Montgomery M, Arnett J. Erikson's young adulthood and emerging adulthood today. J Child Youth Care Work. 2020;25:206-13.
10. Hatfield J, Fernandes R. The role of risk-propensity in the risky driving of younger drivers. Accid Anal Prev. 2009;41:25-35.
11. Sharma B, Stepanovic L, Cheerasarn S, Kiernan SR, Agriantonis G, Bhatia ND, et al. Correlating various clinical outcomes associated with motor vehicle collision-related trauma. Healthcare (Basel). 2025;13:2314.
12. Takagi K, Saitoh D, Nakagawa K, Inoue H, Takeuchi H, Takyu H, et al. Alcohol-intoxicated patients with blunt trauma and head injuries have better outcomes than sober patients. Cureus. 2024;16:e63044.

13. Jurkovich GJ, Rivara FP, Gurney JG, Seguin D, Fligner CL, Copass M. Effects of alcohol intoxication on the initial assessment of trauma patients. Ann Emerg Med. 1992;21:704-8.
14. Beydoun H, Teel A, Crowder C, Khanal S, Lo BM. Past blood alcohol concentration and injury in trauma center: propensity scoring. J Emerg Med. 2014;47:387-94.
15. Göksu E, Çete Y, Kanalıcı H, Kılıçaslan İ. Demographic and clinical properties of patients presenting with traffic accidents and its association with blood alcohol concentration. Turk J Emerg Med. 2008;8:26-31.
16. Scott-Parker B, Watson B, King MJ, Hyde MK. "I drove after drinking alcohol" and other risky driving behaviours reported by young novice drivers. Accid Anal Prev. 2014;70:65-73.
17. Åkerstedt T, Kecklund G. Age, gender and early morning highway accidents. J Sleep Res. 2001;10:105-10.
18. Hsieh CH, Su LT, Wang YC, Fu CY, Lo HC, Lin CH. Does alcohol intoxication protect patients from severe injury and reduce hospital mortality? Am Surg. 2013;79:1289-94.
19. Ahmed N, Greenberg P. Assessing the impact of blood alcohol concentration on the rate of in-hospital mortality following traumatic motor vehicle crash injury: a matched analysis of the National Trauma Data Bank. Injury. 2019;50:33-8.
20. Brockamp T, Böhmer A, Lefering R, Bouillon B, Wafaisade A, Mutschler M, et al. Alcohol and trauma: the influence of blood alcohol levels on the severity of injuries and outcome of trauma patients - a retrospective analysis of 6268 patients of the TraumaRegister DGU®. Scand J Trauma Resusc Emerg Med. 2021;29:101.
21. Yaghoubian A, Kaji A, Putnam B, De Virgilio N, De Virgilio C. Elevated blood alcohol level may be protective of trauma patient mortality. Am Surg. 2009;75:950-3.
22. Hoonpongsimanont W, Ghanem G, Sahota P, Arif A, Barrios C, Saadat S, et al. Association of blood alcohol and alcohol use disorders with emergency department disposition of trauma patients. West J Emerg Med. 2022;23:158-65.
23. Swearingen A, Ghaemmaghami V, Loftus T, Swearingen CJ, Salisbury H, Gerkin RD, et al. Extreme blood alcohol level is associated with increased resource use in trauma patients. Am Surg. 2010;76:20-4.
24. Rundhaug NP, Moen KG, Skandsen T, Schirmer-Mikalsen K, Lund SB, Hara S, et al. Moderate and severe traumatic brain injury: effect of blood alcohol concentration on Glasgow Coma Scale score and relation to computed tomography findings. J Neurosurg. 2015;122:211-8.
25. Hoonpongsimanont W, Ghanem G, Saadat S, Nguyen M, Louis C, Sahota PK, et al. Correlation between alcohol use disorders, blood alcohol content, and length of stay in trauma patients. J Emerg Trauma Shock. 2021;14:42-7.
26. Kelley KC, Salen P, Wojda TR, Hasani AZ, Luster J, Seoane J, et al. Impact of blood alcohol concentration on hematologic and serum chemistry parameters in trauma patients. Int J Crit Illn Inj Sci. 2021;11:18-24.
27. Leijdesdorff HA, Legué J, Krijnen P, Rhemrev S, Kleinveld S, Schipper IB. Traumatic brain injury and alcohol intoxication: effects on injury patterns and short-term outcome. Eur J Trauma Emerg Surg. 2021;47:2065-72.
28. Riuttanen A, Jäntti SJ, Mattila VM. Alcohol use in severely injured trauma patients. Sci Rep. 2020;10:17914.
29. Afshar M, Netzer G, Salisbury-Afshar E, Murthi S, Smith GS. Injured patients with very high blood alcohol concentrations. Injury. 2016;47:83-8.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Case Report
DOI: 10.4274/eajem.galenos.2026.82356
Eurasian J Emerg Med. 2026;25: 529-34

High-risk Pulmonary Embolism Presenting with Inferior ST-segment Elevation and Atrioventricular Block: A Multimodality Case Report

Denizhan Karaçimen, Ahmet Berk Duman, Yunus Koç, Yalçın Dalgıç

University of Health Sciences Türkiye, Kocaeli City Hospital, Clinic of Cardiology, Kocaeli, Türkiye

Abstract

Pulmonary embolism (PE) can rarely mimic an inferior ST-segment elevation myocardial infarction with atrioventricular (AV) block. A 66-year-old woman with morbid obesity collapsed after sudden chest pain and dyspnea. After prehospital and recurrent in-hospital cardiac arrests, she remained intubated and in profound shock. The initial electrocardiogram showed inferior ST elevation, reciprocal changes, and second-degree AV block (Mobitz type 1). Emergency coronary angiography, performed after heparin and cangrelor administration, did not demonstrate an obstructive culprit. Subsequently, bedside echocardiographic reassessment revealed a central echogenic mass at the main pulmonary artery bifurcation despite initially limited signs of right ventricular pressure overload. Computed tomography pulmonary angiography was unsafe because of hemodynamic instability; catheter pulmonary angiography confirmed central obstruction and complete left pulmonary artery occlusion. Mechanical fragmentation using a pigtail catheter was unsuccessful; aspiration was not performed. Following individualized assessment of marked bleeding risk, 20 mg alteplase was administered locally as a non-standard salvage regimen (10 mg into the right pulmonary artery, followed by 10 mg into the left, over approximately 2 minutes per side), without a subsequent infusion. Refractory shock persisted, vascular access site bleeding developed, and the patient subsequently died. Bedside reassessment after culprit-negative coronary angiography identified the cause of refractory shock; the optimal reperfusion approach remains uncertain.

Keywords: Atrioventricular block, case report, echocardiography, pulmonary embolism, ST-segment elevation myocardial infarction

Introduction

Acute pulmonary embolism (PE) can mimic ST-segment elevation myocardial infarction (STEMI), usually with anterior rather than inferior ST elevation (1). Inferior STEMI-like PE with culprit-negative coronary angiography has been reported (2,3), as have PE-associated atrioventricular (AV) conduction disturbances (4,5). Direct visualization of central pulmonary arterial thrombus by transthoracic echocardiography (TTE) is uncommon (6,7). Inferior ST elevation with conduction block can be indistinguishable from a coronary emergency during initial resuscitation.

After cardiac arrest, an electrocardiogram (ECG) that appears coronary may prompt urgent angiography before PE is suspected. We describe the diagnostic transition from inferior ST-elevation and AV block to negative coronary angiography, the bedside detection of a central pulmonary arterial mass, the confirmatory catheter angiography, and the difficult reperfusion decisions. The timing of the transition is central to its clinical teaching point.

Case Report

A 66-year-old woman (body mass index of 41 kg/m^2) without documented cardiovascular disease, prior thromboembolism, malignancy, or recent surgery collapsed following the abrupt onset of chest pain and dyspnea. Prehospital cardiopulmonary resuscitation (CPR) lasted approximately 15–20 minutes before the return of spontaneous circulation (ROSC; ~21:50). Multiple rib fractures were observed. At the referring hospital (~22:10), blood gases showed pH 6.98, bicarbonate 8.9 mmol/L, lactate 11.8 mmol/L, PaO_2 88.2 mmHg, and $PaCO_2$ 43.0 mmHg; the ECG showed inferior ST elevation. She was intubated and transferred

Corresponding Author: Ahmet Berk Duman, MD, University of Health Sciences Türkiye, Kocaeli City Hospital, Clinic of Cardiology, Kocaeli, Türkiye
E-mail: dumanahmetberk@gmail.com **ORCID ID:** orcid.org/0009-0008-1678-4272

Cite this article as: Karaçimen D, Duman AB, Koç Y, Dalgıç Y. High-risk pulmonary embolism presenting with inferior ST-segment elevation and atrioventricular block: a multimodality case report. Eurasian J Emerg Med. 2026;25: 529-34

Received: 22.08.2026
Accepted: 05.11.2026
Published: 16.11.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

while receiving vasoactive support, approximately 1 hour after presentation to the referring hospital. A recurrent arrest shortly after arrival required ~10 minutes of CPR (Table 1).

After ROSC, she remained intubated, ventilated, and hypotensive (approximately 70/40-60/30 mmHg) despite an infusion of noradrenaline at 0.06 µg/kg/min. The repeat blood gas showed pH 7.19, bicarbonate 18.1 mmol/L (measured after sodium bicarbonate), lactate 12.58 mmol/L, PaO_2 62.2 mmHg, $PaCO_2$ 48.6 mmHg, and arterial oxygen saturation 84.4%. The referring hospital ECG showed sinus rhythm with second-degree AV block (Mobitz type 1 [Wenckebach], 4:3 conduction); a ventricular rate of approximately 44 beats/min; inferior ST elevation (II, III, aVF) with reciprocal ST depression in leads I and aVL; and subtle ST elevation in V1 (Figure 1A). The STEMI pathway was activated. The inferior ECG pattern and persistent shock initially made acute coronary occlusion a plausible working diagnosis.

For suspected inferior STEMI, 10,000 IU unfractionated heparin and intravenous cangrelor were administered at ~23:30, before femoral coronary angiography (~23:35), which found no obstructive culprit (Figure 2A, 2B). The coronary findings did not explain the severity of shock. High-sensitivity troponin I was 904 ng/L (reference <26), BNP was 3000 pg/mL, and D-dimer was 30.9 µg/mL; initial hemoglobin was 13.7 g/dL and platelets were 110×10^9/L.

Persistent unexplained shock after culprit-negative angiography prompted immediate bedside TTE. Left ventricular systolic function was preserved, without prominent early right ventricular (RV) dilatation or septal flattening. Parasternal short-axis imaging at the great-vessel level revealed a non-mobile echogenic mass (~3.0×1.3 cm) at the main pulmonary artery bifurcation (Figure 3A, Supplementary Video 1), suggestive of a saddle-like thrombus. PE became the leading diagnosis. Initial tricuspid regurgitation peak velocity was ~2.5 m/s, and the inferior vena cava diameter was 1.3 cm, without visible respiratory collapse. Estimated pulmonary artery systolic pressure was ~25 mmHg, a value considered unreliable during low-flow shock. Computed tomography pulmonary angiography (CTPA) was deferred because transport was unsafe. The bedside mass was considered a trigger for confirmatory PE-directed imaging, but not as definitive proof.

For urgent reperfusion, catheter pulmonary angiography was performed at ~00:30 via femoral venous access with a 5-Fr pigtail catheter into the main pulmonary artery, after a risk discussion and written procedural consent from next of kin. This strategy permitted confirmation and potential treatment without an additional diagnostic transfer. Anteroposterior (AP 0°/0°) and left anterior oblique (LAO 30°/0°) views showed a central bifurcation filling defect, a predominantly left-sided thrombus, and complete occlusion of the left pulmonary artery (Figure 3B). Angiography did not establish extension into both main branches. Despite the saddle-like TTE appearance, a definite

Table 1. Approximate clinical timeline

Time*	Clinical event and diagnostic/therapeutic action
Symptom onset	Sudden chest pain, dyspnea, and collapse at home; exact onset time unknown.
~21:50	First ROSC after approximately 15-20 min of prehospital CPR.
~22:10	Referring-hospital assessment; initial ECG and arterial blood gases.
~1 h after referring-hospital presentation	Transfer to treating center; recurrent arrest shortly after arrival and ~10 min further CPR.
~23:30	Working diagnosis of inferior STEMI; unfractionated heparin and intravenous cangrelor.
~23:35	Emergency coronary angiography: no obstructive culprit lesion.
~23:45-00:00	Immediate bedside reassessment: central pulmonary arterial mass visualized on TTE; CTPA deferred due to instability.
~00:30	Pulmonary angiography in AP and LAO 30° projections: complete left pulmonary artery occlusion.
After pulmonary angiography	Unsuccessful 5-Fr pigtail mechanical fragmentation; no aspiration performed; informed discussion/consent for local alteplase.
~00:50	Local alteplase: right PA 10 mg over ~2 min, then left PA 10 mg over ~2 min (20 mg total); no further infusion.
~01:30	Follow-up ECG and TTE: resolved AV block and evolving RV failure.
~03:30	Vascular access-site bleeding and hemoglobin decline; O RhD-negative red cells requested but not available before death.
~06:00	Death during persistent refractory shock.

*All clock times are approximate and relative to the presenting episode.
CPR: Cardiopulmonary resuscitation, CTPA: Computed tomography pulmonary angiography, ECG: Electrocardiogram, ROSC: Return of spontaneous circulation, STEMI: ST-segment elevation myocardial infarction, TTE: Transthoracic echocardiography

anatomical saddle classification was avoided. Recurrent arrest and refractory vasopressor-dependent shock supported high-risk PE (AHA/ACC acute PE category E2) (8).

After heparin/cangrelor and before alteplase, INR was 2.38, fibrinogen was 0.64 g/L, and aPTT was unmeasurable because no clot formed; the team knew these results before treatment. Full-dose systemic fibrinolysis was considered but ultimately not selected after individualized assessment of coagulopathy, traumatic CPR with rib fractures, antithrombotic exposure, and femoral arterial/venous access. Surgical embolectomy and extracorporeal membrane oxygenation were unavailable. Mechanical fragmentation using the 5-Fr pigtail failed to achieve adequate reperfusion; the technical reason remained undetermined. No aspiration thrombectomy was performed. The hemorrhagic risks were considered cumulatively; no single risk factor was presented as an absolute contraindication. In

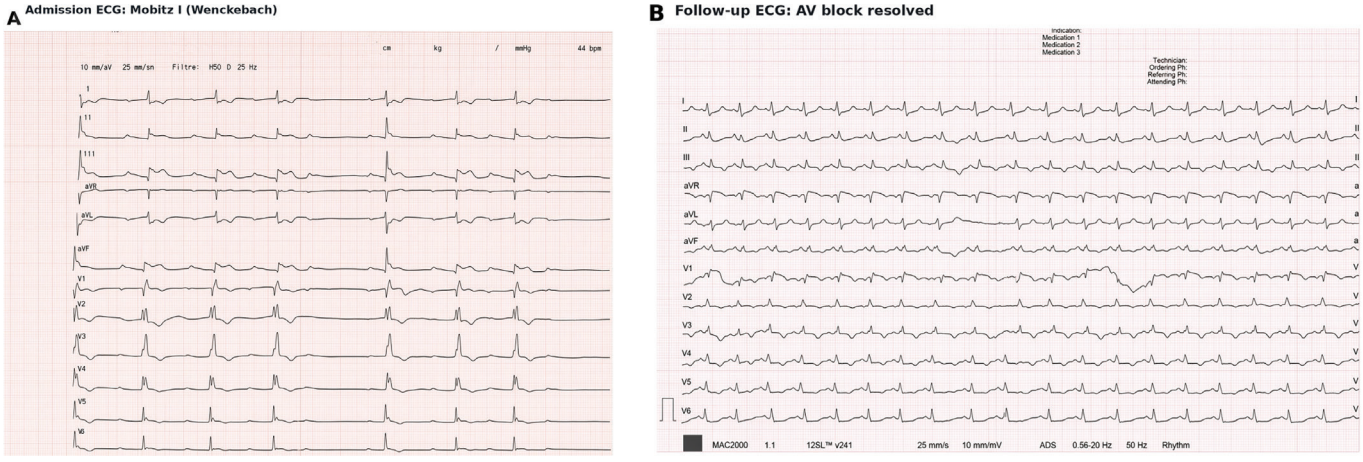


Figure 1. Electrocardiographic evolution. (A) Initial 12-lead ECG obtained at the referring hospital showing second-degree AV block, Mobitz type 1 (Wenckebach), with 4:3 AV conduction and a ventricular rate of approximately 44 beats/min, inferior ST-segment elevation in leads II, III and aVF with reciprocal ST depression in I and aVL, and subtle ST elevation in V1. (B) Follow-up ECG during ongoing reperfusion and hemodynamic management demonstrating sinus tachycardia, resolution of AV block, S1Q3T3, new right bundle branch block, and T-wave inversion in V1-V4

ECG: Electrocardiogram, AV: Atrioventricular

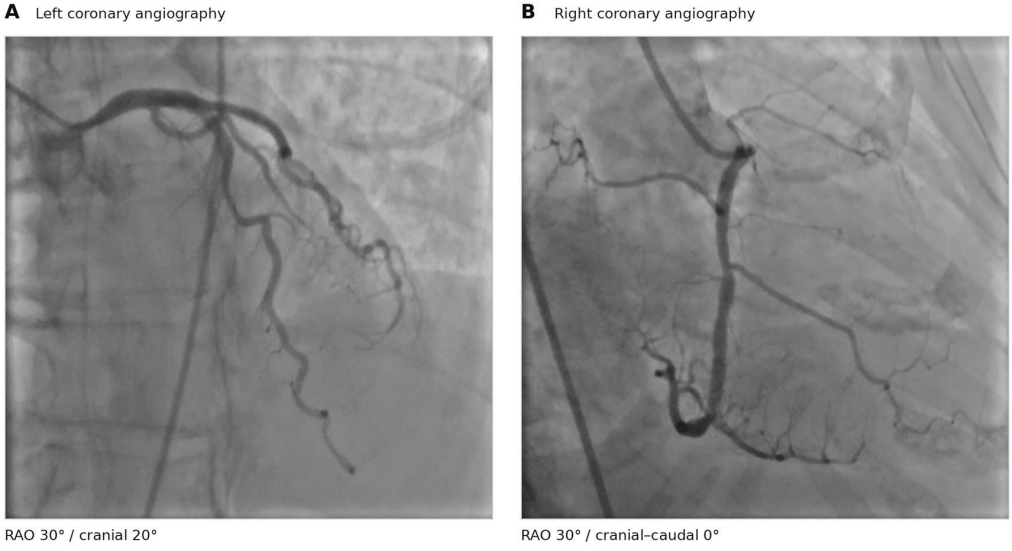


Figure 2. Emergency coronary angiography. (A) Left coronary angiography in a right anterior oblique cranial projection (RAO 30°, cranial 20°) demonstrating no obstructive culprit lesion. (B) Right coronary angiography in a right anterior oblique projection (RAO 30°, cranial/caudal 0°) demonstrating no obstructive culprit lesion. The displayed epicardial vessels are patent without an angiographically identified acute culprit

RAO: Right anterior oblique

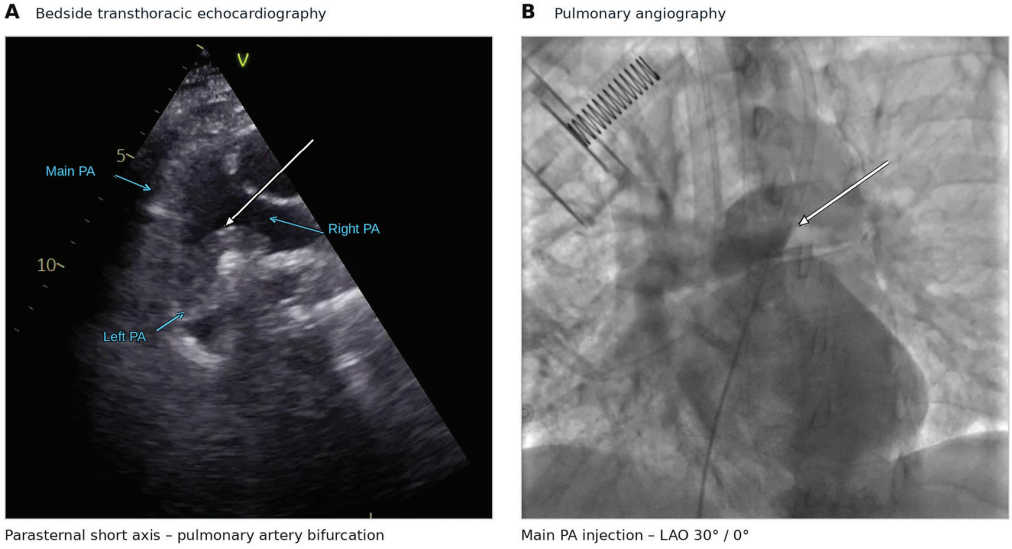


Figure 3. Multimodality identification of central pulmonary embolism. (A) Bedside transthoracic echocardiography in the parasternal short-axis view at the great-vessel level showing a non-mobile echogenic mass (approximately 3.0×1.3 cm) at the main pulmonary artery bifurcation with a saddle-like appearance. The white arrow marks the echogenic mass; blue arrows label the main and branch pulmonary arteries. (B) Catheter pulmonary angiography after contrast injection into the main pulmonary artery, shown in the LAO 30° projection (no cranial/caudal angulation), demonstrating central filling defect and complete left pulmonary artery occlusion; an AP 0°/0° view was also acquired. Angiography did not establish definite bilateral saddle extension

LAO: Left anterior oblique, AP: Anteroposterior, PA: Pulmonary artery

the setting of refractory shock, the team discussed the risk of major bleeding and potential benefits with the next of kin, and obtained written consent. At ~00:50, 20 mg of alteplase was delivered locally through the pigtail catheter: 10 mg into the right pulmonary artery over ~2 minutes, followed by 10 mg into the left pulmonary artery over ~2 minutes. No further alteplase bolus or infusion was given. This was a non-standard salvage regimen. The exact distal catheter-tip position during each injection was not separately recorded.

Hypotension worsened despite escalation of noradrenaline from 0.06 to 1.5 µg/kg/min and addition of dopamine (~2.8 µg/kg/min). At ~01:30, AV block had resolved during ongoing management; the ECG showed sinus tachycardia, S1Q3T3, new right bundle branch block, and anterior T-wave inversion (Figure 1B). Repeat TTE showed RV dilatation, septal flattening with a D-shaped left ventricle, significant tricuspid regurgitation, a pulmonary acceleration time of ~61 ms, and an inferior vena cava diameter of 1.9 cm without respiratory collapse (Figure 4A-4C). Estimated pulmonary artery systolic pressure was ~30 mmHg but was unreliable in low-flow RV failure. At ~03:30, oozing was noted from the femoral and intravenous access sites, and local dressings were applied. Hemoglobin decreased from 13.7 g/dL to 10.6 g/dL and then to 6.5 g/dL. Packed red cells were requested; however, O RhD-negative units were unavailable in the hospital blood bank and had not arrived before death at ~06:00. The source of the hemoglobin decline was not fully established, and the contribution of bleeding to death during refractory shock remains uncertain.

Discussion

Inferior STEMI-like PE with culprit-negative coronary angiography (2,3), PE-associated AV block (4,5), and central thrombus detected by TTE (6,7) have each been reported. Our case emphasizes their diagnostic sequence after arrest: an apparent coronary event, negative angiography, an immediate, focused TTE, and confirmatory pulmonary angiography. Direct thrombus visualization was especially useful because conventional RV strain findings were initially limited.

RV pressure overload or ischemia, supply-demand mismatch, hypoxemia, and post-arrest ischemia may contribute to ST changes (1-3). A transient coronary event or vasospasm cannot be excluded on the basis of a later culprit-negative angiogram. AV block may reflect RV ischemia, autonomic effects, or post-arrest physiology; its resolution during treatment does not establish a causal effect of alteplase. This was a STEMI-like presentation with confirmed PE; it does not prove a single mechanism. The serial ECGs document the evolution but cannot identify a unique electrophysiological pathway.

Echocardiography alone cannot confirm or exclude PE because indirect RV findings lack specificity (8). Here, direct bifurcation-mass visualization, despite limited early RV signs, prompted

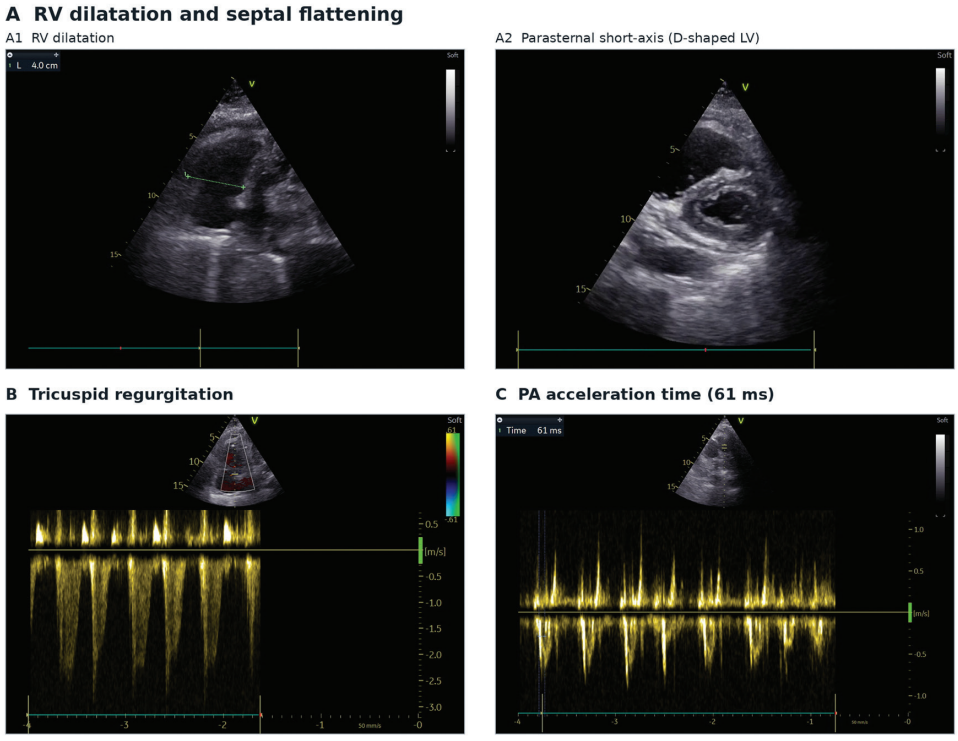


Figure 4. Echocardiographic markers of right ventricular (RV) failure. (A) Repeat transthoracic echocardiography showing RV dilatation (A1) and interventricular septal flattening with a D-shaped left ventricle in the parasternal short-axis view (A2). (B) Doppler echocardiography demonstrating tricuspid regurgitation. (C) Pulmonary artery Doppler showing a shortened acceleration time of approximately 61 ms. The estimated pulmonary artery systolic pressure (~30 mmHg) was interpreted cautiously in the setting of severe low-flow RV failure

PA: Pulmonary artery

catheter angiography. The saddle-like TTE appearance was not definitive; angiography confirmed central obstruction and left pulmonary artery occlusion but did not show unequivocal bilateral branch extension. Similar bedside ultrasound redirection has been described (10). Angiography confirmed the obstruction but did not resolve all anatomical uncertainties suggested by TTE.

Advanced reperfusion in PE with cardiopulmonary failure depends on the hemorrhagic risk and on local resources (8,9). Known pre-alteplase coagulation abnormalities, traumatic CPR, antithrombotic exposure, and femoral access increased bleeding risk, but did not individually constitute absolute contraindications to systemic thrombolysis. With surgery and extracorporeal support unavailable, and fragmentation unsuccessful, the team selected rapid local administration of alteplase after a documented risk-benefit discussion and written consent. The total alteplase dose (20 mg) was administered over approximately 4 minutes without further infusion; however, this is not an established catheter-directed thrombolysis protocol. Subsequent bleeding had multiple contributing factors and no single identifiable cause. The effect of this regimen on survival or bleeding cannot be determined from a single case.

Limitations include unavailability of CTPA, invasive pulmonary pressure measurements, and autopsy. RV/LV ratio, TAPSE, and RV S′ could not be reliably retrieved from emergency images; therefore, RV enlargement was assessed qualitatively. Access-site oozing did not account for the full extent of the hemoglobin decline; transfusion was not begun before death. Neither the independent contribution of hemorrhage nor the question of whether earlier TTE or a different reperfusion strategy would have changed survival can be determined from a single case.

Conclusion

High-risk PE may mimic inferior STEMI with AV block. After culprit-negative coronary angiography in unstable post-arrest shock, bedside TTE may reveal central PE and guide confirmatory imaging when CTPA is unsafe. TTE cannot independently confirm PE or determine a preferred reperfusion strategy.

Ethics

Informed Consent: Written informed consent for publication of this case report and accompanying electrocardiogram, echocardiographic, coronary angiographic, pulmonary angiographic, and video material was obtained from the patient's

next of kin. The treating team discussed the substantial risks and potential benefits of pulmonary embolism-directed invasive treatment and local thrombolysis with the patient's next of kin. Written consent was obtained before treatment.

Footnotes

Authorship Contributions

Surgical and Medical Practices - Concept - Design - Data Collection or Processing - Analysis or Interpretation - Literature Search - Writing: All authors contributed equally to this study.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

Acknowledgements

The authors thank Samet Solmaz, MD, Clinic of Cardiology, Kocaeli City Hospital, for assistance with the clinical care of the patient.

Reporting Guideline: This report was prepared in accordance with CARE recommendations; a completed CARE checklist is supplied separately. The patient's perspective was unavailable because she remained intubated after the arrest and died during the index admission.

Declaration Regarding the Use of AI and AI-Assisted Technologies: During manuscript preparation, ChatGPT (OpenAI) was used for language editing, journal-specific restructuring, and editorial critique. The authors reviewed and verified the final text and take full responsibility for its content.

Supplementary Video 1. Parasternal short-axis cine at the level of the great vessels demonstrating an echogenic mass at the main pulmonary artery bifurcation

Available from: https://youtu.be/HZ05crcval4

References

1. Villablanca PA, Vlismas PP, Aleksandrovich T, Omondi A, Gupta T, Briceno DF, et al. Case report and systematic review of pulmonary embolism mimicking ST-elevation myocardial infarction. Vascular. 2019;27:90-7.
2. Emren SV, Arikan ME, Senoz O, Varis E, Akan E. Acute pulmonary embolism mimicking inferior myocardial infarction. Turk Kardiyol Dern Ars. 2014;42:290-3.
3. Mistry A, Natarajan N, Hussain S, Vali Z. Unusual presentation of acute pulmonary embolus presenting with inferior ST elevation. BMJ Case Rep. 2018;2018:bcr2018226063.
4. Akinboboye OO, Brown EJ Jr, Queirroz R, Cusi VP, Horowitz LF, Jonas EA, et al. Recurrent pulmonary embolism with second-degree atrioventricular block and near syncope. Am Heart J. 1993;126:731-2.
5. Ma M, Liang S, He Y, Wang H. Case report: The presence of third-degree atrioventricular block caused by pulmonary embolism masquerading as acute ST-segment elevation myocardial infarction. Front Cardiovasc Med. 2023;10:1013727.
6. Yuriditsky E, Horowitz JM, Taslakian B, Saric M. Saddle pulmonary embolism detected by transthoracic echocardiography in a patient with suspected myocardial infarction. CASE (Phila). 2024;8:54-7.
7. Lipa AJ, Domanovits H, Schwameis M. Saddle pulmonary embolism detected by transthoracic echocardiography-a case report. Wien Klin Wochenschr. 2026.
8. Writing Committee Members; Creager MA, Barnes GD, Giri J, Mukherjee D, Jones WS, Burnett AE, et al. 2026 AHA/ACC/ACCP/ACEP/CHEST/SCAI/SHM/SIR/SVM/SVN Guideline for the evaluation and management of acute pulmonary embolism in adults: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 2026;153:e977-1051.
9. Pruszczyk P, Klok FA, Kucher N, Roik M, Meneveau N, Sharp ASP, et al. Percutaneous treatment options for acute pulmonary embolism: a clinical consensus statement by the ESC Working Group on pulmonary circulation and right ventricular function and the European Association of Percutaneous Cardiovascular Interventions. EuroIntervention. 2022;18:e623-38.
10. Shy BD, Gutierrez A, Strayer RJ. Bedside ultrasound to evaluate pulmonary embolism masquerading as ST elevation myocardial infarction (STEMI). J Emerg Med. 2015;49:703-4.e1.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2026.76735
Eurasian J Emerg Med. 2026;25: 535-42

Attenuated Composite Inflammatory Indices in Polymerase Chain Reaction-confirmed Pediatric Influenza Presenting to the Emergency Department

Muhammed Kamiloğlu[1], Keziban Uçar Karabulut[1], Seyit Ali Karagöz[1], Ozan Özelbaykal[2]

[1]Başkent University Konya Application and Research Center, Department of Emergency Medicine, Konya, Türkiye
[2]Konya Numune Hospital, Clinic of Emergency Medicine, Konya, Türkiye

Abstract

Aim: To compare inflammatory indices derived from blood counts between influenza positive and negative children evaluated in the emergency department and examine the association of pan-immune-inflammation value (PIV) with influenza positivity.

Materials and Methods: This retrospective study included 903 children with upper respiratory tract symptoms who underwent blood counts, C-reactive protein (CRP) measurement, and influenza polymerase chain reaction testing at the same emergency visit. Group comparisons and receiver operating characteristic analyses were conducted using the full sample. An exploratory regression within overlapping ages (n=783) was performed, adjusting for age and CRP, using restricted cubic splines.

Results: Influenza was positive in 366 children (40.5%). The systemic immune-inflammation index (SII), the systemic inflammation response index (SIRI), and the PIV were lower in positive children (all $p<0.001$). Areas under the curve were 0.575, 0.619, and 0.648, respectively, with lower values indicating positivity. Across overlapping age groups, the adjusted odds ratio per PIV doubling was 0.746 (95% confidence interval, 0.672-0.828). Bootstrap-corrected model discrimination increased from 0.700 to 0.731 after adding PIV.

Conclusion: Lower composite indices were associated with influenza positivity in this selected sample. Limited discrimination and sampling constraints preclude claims of validated clinical diagnostic utility.

Keywords: Influenza, children, emergency department, blood cell count, inflammatory indices

Introduction

Influenza remains an important cause of childhood respiratory illness and may lead to serious complications (1,2). Its symptoms overlap with those of other respiratory infections; therefore, virologic testing, including polymerase chain reaction (PCR), remains central to establishing influenza status (3). Complete blood count (CBC) findings may describe the accompanying cellular pattern, but an association with infection does not establish that a marker improves diagnosis or patient care.

Previous pediatric studies have assessed leukocyte and platelet counts and ratios, such as the neutrophil-to-lymphocyte ratio (NLR) and the lymphocyte-to-monocyte ratio, with their performance varying by age, comparator population, and reference test (4-6). Hematologic markers have also been examined in predominantly adult respiratory infection cohorts (7). The systemic immune-inflammation index (SII), systemic inflammation response index (SIRI), and pan-immune-inflammation value (PIV) are based on several cell populations. Evidence from pneumonia requiring intensive care, comparisons of influenza and Coronavirus disease-2019 (COVID-19) pneumonia, and pediatric infections shows that the clinical setting and outcome must be considered when interpreting such measures (8-10).

A recent pediatric influenza study evaluated an aggregate index of systemic inflammation, using the same cell-count formula as PIV, to predict pulmonary complications among hospitalized

Corresponding Author: Muhammed Kamiloğlu, MD, Başkent University Konya Application and Research Center, Department of Emergency Medicine, Konya, Türkiye
E-mail: mkamiloglu@baskent.edu.tr **ORCID ID:** orcid.org/0000-0003-0317-018X

Cite this article as: Kamiloğlu M, Uçar Karabulut K, Karagöz SA, Özelbaykal O. Attenuated composite inflammatory indices in polymerase chain reaction-confirmed pediatric influenza presenting to the emergency department. Eurasian J Emerg Med. 2026;25: 535-42

Received: 11.09.2026
Accepted: 30.10.2026
Published: 18.11.2026

©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House. Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

children (11). That prognostic question differs from a comparison of influenza-positive and influenza-negative children presenting with upper respiratory tract symptoms. In a selected emergency department (ED) cohort, we examined differences in SII, SIRI, and PIV by influenza PCR status, their correlations with C-reactive protein (CRP), and the association of PIV with influenza positivity after adjustment for age and CRP. The analyses were exploratory; no index or diagnostic threshold was prospectively prespecified as a primary endpoint.

Materials and Methods

Study Design and Population

This retrospective observational study included children aged <18 years who were evaluated for upper respiratory tract symptoms in the ED of a secondary care hospital between January 1, 2023, and January 1, 2025. Demographic, laboratory, influenza test, and medical history data were extracted from the ED hospital information management system. The study was approved by the Başkent University Institutional Review Board (protocol no: KA25/476, date: 13.01.2026), which waived the requirement for informed consent because of its retrospective, anonymized design. Reporting followed the STROBE framework (12).

Patient Selection

Upper respiratory tract symptoms were the clinical indications for testing. Eligibility required CBC, CRP, and influenza results obtained at the same ED encounter, as well as sufficient medical history to assess exclusion criteria. For children who attended more than once, the first eligible visit with the required tests was retained, and subsequent visits were excluded from the analysis. The screened denominator, 1,208, represents unique children after removal of repeat encounters, rather than all ED visits or all children with respiratory symptoms.

Children with a malignancy, those receiving immunosuppressive treatment, those with an autoimmune disease, or those missing CBC, CRP, or influenza results, or those with inadequate medical history were excluded. A positive throat or urine culture and/or a procalcitonin (PCT) level >0.05 ng/mL within one week of the index visit constituted an additional operational exclusion criterion. Cultures and PCT were performed only in selected patients. A positive culture led to exclusion even when contamination was suspected. This rule was not equivalent to confirmation of bacterial coinfection for every excluded child, and bacterial coinfection was not systematically assessed throughout the cohort.

Virologic Testing and Laboratory Variables

Nasal swab specimens were used for PCR testing targeting influenza A and influenza B. Children positive for either target were included in the influenza-positive group; dual A+B positivity was counted once. The comparator group had negative influenza A/B results. Alternative etiologic diagnoses and other pathogen test results were not available in the analyzed dataset; therefore, a negative influenza result was not considered evidence of a specific alternative infection. CBC, CRP, and influenza samples were collected during the same ED encounter; symptom duration and precise temporal order were not available in the analyzed dataset.

Absolute leukocyte, neutrophil, lymphocyte, monocyte, and platelet counts were expressed in $10^3/\mu L$; hemoglobin in g/dL; red cell distribution width (RDW) in percent; and CRP in mg/L. Indices were recalculated as follows: NLR=neutrophils/lymphocytes; MLR=monocytes/lymphocytes; PLR=platelets/lymphocytes; Hb/RDW=hemoglobin/RDW; SII=platelets×neutrophils/lymphocytes; SIRI=neutrophils×monocytes/lymphocytes; and PIV=platelets×neutrophils×monocytes/lymphocytes. The composite indices were computed using the stated numerical cell-count units; they should not all be regarded as dimensionless.

Statistical Analysis

Data and statistical results were checked using IBM SPSS statistics version 27.0 (IBM Corp., Armonk, NY, USA). Statistical analyses were performed in Python 3.12.14 using NumPy 2.3.5, pandas 2.2.3, SciPy 1.17.0, and scikit-learn 1.8.0. Results from SPSS were consistent with the corresponding Python outputs for the analyses checked. Influenza positivity was coded as 1 and negativity as 0. Shapiro-Wilk tests were used to assess the distributions of continuous variables. Non-normally distributed variables were summarized by median, 25th and 75th percentiles, and range. Quartiles were calculated using linear interpolation. Two-sided Mann-Whitney U tests were used to compare continuous distributions, and Pearson's chi-square test was used to compare sex. Raw p values were reported, with an additional Holm adjustment applied across the seven index comparisons. Spearman's correlations were used to assess the associations of SII, SIRI, and PIV with CRP. Variables for analysis were complete in the final sample; no imputation was performed.

Single-index receiver operating characteristic (ROC) analysis included all 903 children; lower index values indicated influenza positivity. Areas under the curve (AUCs) and 95% confidence intervals (CIs) were calculated using the non-parametric DeLong variance (13). Youden index thresholds and their apparent operating characteristics were calculated in the same sample as exploratory estimates and were not considered validated clinical decision thresholds.

The full-cohort logistic model was assessed for logit linearity using Box-Tidwell-type interaction terms and for calibration

using a decile Hosmer-Lemeshow test. Concerns regarding logit linearity and calibration, together with limited age overlap between the groups, prompted additional exploratory analyses. Descriptive comparisons and single-index ROC analyses included all 903 children. Adjusted analyses were restricted post hoc to the observed common age interval of 2.25-9.9167 years, comprising 783 children. This restriction limits extrapolation; it does not eliminate selection bias or establish applicability to older children.

Age and $\log_2$(CRP) were modeled with restricted cubic splines using four knots at the empirical 5th, 35th, 65th, and 95th percentiles (14). PIV was $\log_2$-transformed, so its odds ratio (OR) represents the change associated with a doubling. Age and CRP were retained for clinical relevance. PIV was chosen during exploratory analysis as a representative composite measure incorporating neutrophils, monocytes, platelets, and lymphocytes; this selection was informed by the observed results. Mathematically overlapping composite indices were not entered simultaneously. Univariable significance screening, stepwise elimination, and penalized variable selection were not used. Global likelihood-ratio tests assessed the adjusted and unadjusted terms. The spline effects of age and CRP were not summarized by a single, constant OR.

The age-and-CRP model was compared with the model adding PIV using a likelihood-ratio test; the change in AUC was reported descriptively. In-sample DeLong significance tests were not used for these nested fitted models (15). Internal validation used 500 bootstrap resamples (seed 20261014), re-estimating knots and refitting the specified models in each resample. Optimism-corrected AUC, Brier score, calibration slope, and intercept were estimated (16). This validation was conditional on the specified model structure and did not account for the selection of PIV from among the candidate indices.

Additional non-linear PIV terms, sex adjustment, restriction to age <10 years, exclusion of the upper 1% of PIV values, and exclusion of dual A+B positive records were examined. Term-level generalized variance inflation factors (GVIFs), leverage, approximate Cook's distance, and Pearson residuals were inspected. GVIF >5 was used as a descriptive flag for potential collinearity, alongside term degrees of freedom and $GVIF^{(1/[2\times df])}$; it was not used as an automatic exclusion criterion. No observations were removed from the main analyses on the grounds of statistical influence. A separate, unpenalized logistic regression solver evaluated the fitted probabilities. Tests were two-sided, with nominal $p<0.05$. Model and sensitivity analyses were exploratory.

Results

Participant Selection and Baseline Characteristics

Of the 1,208 unique children assessed, 305 were excluded: 51 for the culture/PCT rule, 3 for malignancy, 8 for immunosuppressive treatment, and 243 for missing required laboratory results or an adequate medical history. Each excluded child was assigned to one exclusion group; none were excluded because of autoimmune disease. Among the 51 who met the culture/PCT rule, 39 had PCT >0.05 ng/mL, 37 had positive cultures, and 25 had both. Thus, 14 met only the PCT criterion and 12 only the culture criterion; these categories do not establish that the other test was performed and found to be negative. The final sample comprised 903 children (Figure 1).

Influenza was positive in 366 children (40.5%) and negative in 537 children (59.5%). Among positive children, 241 had influenza A alone, 118 had influenza B alone, and 7 had dual A+B positivity. Positive children were older, whereas the proportions of female children were similar (48.9% vs. 49.0%). Leukocyte, neutrophil, lymphocyte, monocyte, platelet, and CRP values were lower, and hemoglobin was higher in positive children (all $p<0.001$); RDW did not differ significantly (Table 1).

Children who tested negative ranged in age from 2.25 to 9.9167 years, whereas children who tested positive ranged in age from

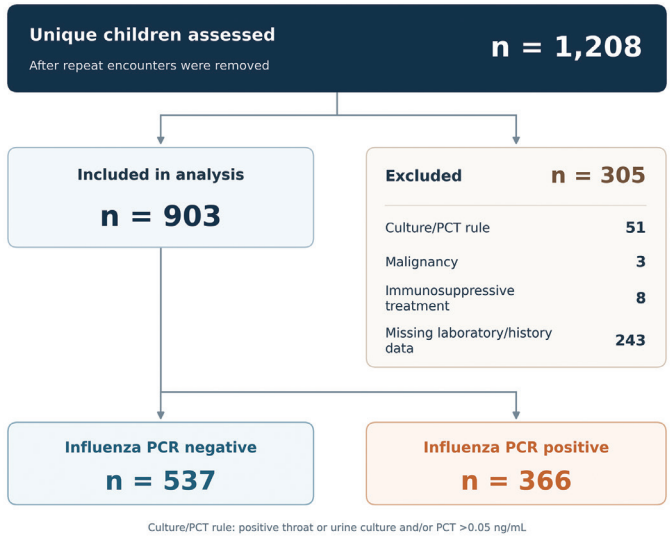


Figure 1. Participant selection. The denominator represents 1,208 unique children after removal of repeat encounters; the first eligible visit with the required tests was retained. Each excluded child was counted in one group. Missing data comprised absent complete blood count, C-reactive protein, influenza results, or adequate medical history. The 51 culture/PCT exclusions were 14 meeting only the PCT criterion, 12 meeting only the culture criterion, and 25 meeting both; "only" does not imply that the other test was performed and negative

PCT: Procalcitonin, PCR: Polymerase chain reaction

Table 1. Baseline characteristics by influenza PCR status			
Variable	**Negative (n=537)**	**Positive (n=366)**	**p**
Age (years)	6.08 (4.08-8.00) [2.25-9.92]	7.50 (5.27-10.48) [1.00-17.92]	<0.001
Female sex, n (%)	263 (49.0)	179 (48.9)	0.984
Leukocytes (10^3/µL)	10.20 (7.64-13.80) [3.09-33.70]	6.03 (4.73-8.44) [2.08-23.50]	<0.001
Neutrophils (10^3/µL)	6.15 (3.87-9.25) [0.42-25.50]	3.15 (2.42-5.37) [0.28-19.10]	<0.001
Lymphocytes (10^3/µL)	2.51 (1.61-3.77) [0.31-17.20]	1.42 (1.08-2.28) [0.33-8.89]	<0.001
Monocytes (10^3/µL)	1.00 (0.72-1.41) [0.09-3.80]	0.67 (0.56-0.92) [0.07-2.97]	<0.001
Platelets (10^3/µL)	275.0 (226.0-335.0) [107.0-756.0]	230.0 (198.3-273.8) [19.5-595.0]	<0.001
Hemoglobin (g/dL)	12.7 (12.0-13.4) [8.5-19.1]	13.1 (12.5-13.7) [6.9-22.1]	<0.001
RDW (%)	12.5 (11.7-13.4) [9.3-20.3]	12.4 (11.7-13.3) [9.3-21.4]	0.493
CRP (mg/L)	12.1 (1.8-42.6) [0.5-301.6]	4.0 (1.1-11.6) [0.5-140.6]	<0.001

Continuous data are median (25^{th}-75^{th} percentiles), with [minimum-maximum]. Quartiles use linear interpolation. Mann-Whitney U tests compare distributions; sex uses Pearson's chi-square test. All p values are unadjusted.
PCR: Polymerase chain reaction, CRP: C-reactive protein, RDW: Red cell distribution width

1.00 to 17.9167 years. All 537 negative children were <10 years old. In the positive group, 256 (69.9%) were <10 years old, 88 (24.0%) were 10-<15 years old, and 22 (6.0%) were 15-<18 years old. The common age interval contained 783 children, including 246 positive and 537 negative; the 120 positive children outside this interval remained in the full descriptive sample.

Inflammatory Indices and Discrimination

SII, SIRI, and PIV were lower in positive children; the median values were 500.8 vs. 723.6 for SII, 1.44 vs. 2.73 for SIRI, and 336.3 vs. 765.4 for PIV, respectively (all raw p<0.001; Table 2, Figure 2). MLR, PLR, and Hb/RDW were higher, while NLR did not differ significantly. All index comparisons except NLR remained significant after Holm adjustment. CRP correlated positively with SII (rho=0.320), SIRI (rho=0.402), and PIV (rho=0.390; all p<0.001).

With lower values indicating positivity, full-cohort AUCs were 0.575 (95% CI 0.537-0.612) for SII, 0.619 (0.582-0.656) for SIRI, and 0.648 (0.612-0.684) for PIV (Figure 3). These indicated limited discrimination, despite statistical evidence of AUCs above 0.5. For the exploratory PIV threshold ≤413.732, apparent sensitivity was 60.9% and specificity 67.8%. Threshold- and common-age ROC results are presented in Supplementary Tables 1, 2.

Exploratory Adjusted Analyses

In the full-cohort logistic regression model with linear terms for age, CRP, and PIV, the joint Box-Tidwell-type test indicated non-linearity (p<0.001), and the decile Hosmer-Lemeshow test yielded p=0.002. In the sample of 783 children within the common age interval, spline analyses showed non-linear associations for age and CRP (p=0.009 and p<0.001, respectively). In this sample, the unadjusted PIV OR per doubling was 0.726 (95% CI 0.663-0.796), and the adjusted OR was 0.746 (95% CI 0.672-0.828; Table 3). The adjusted likelihood-ratio p value for PIV was <0.001. Additional non-linear PIV terms yielded p=0.105, which does not establish that its logit relationship is exactly linear.

Adding PIV improved the model fit (likelihood-ratio χ^2=32.39, 1 degree of freedom, p<0.001). The apparent AUC increased from 0.708 to 0.740, and the optimism-corrected AUCs were 0.700 and 0.731. Corrected Brier scores were 0.194 and 0.185, and corrected calibration slopes were 0.949 and 0.953, respectively (Table 4). All 500 resamples converged. The bootstrap percentile CI for the PIV OR was 0.669-0.822. These estimates quantify internal sample performance but do not indicate validated clinical benefit.

Term-level GVIFs were 1.132 for age, 1.174 for $\log_2$(CRP), and 1.176 for $\log_2$(PIV). Maximum leverage was 0.042 and maximum approximate Cook's distance was 0.018; 34 observations exceeded the descriptive Cook's threshold 4/n, and 8 had absolute Pearson's residuals >3. All observations were retained. PIV estimates were similar after adjustment for sex and in the specified sensitivity samples (Supplementary Tables 1, 2). Fitted probabilities differed by <0.000001 between the two logistic regression solvers.

Discussion

In this selected ED sample, influenza-positive children had lower SII, SIRI, and PIV than influenza-negative children. The PIV association persisted in an exploratory analysis restricted to common ages, after flexible adjustment for age and CRP. Single-index AUCs were modest, and the internally evaluated gain after adding PIV does not demonstrate improved treatment decisions or outcomes. The contribution provides a comparative description of several composite indices and their components, with an explicit assessment of age overlap and model assumptions, rather than introducing a new index or a validated diagnostic test.

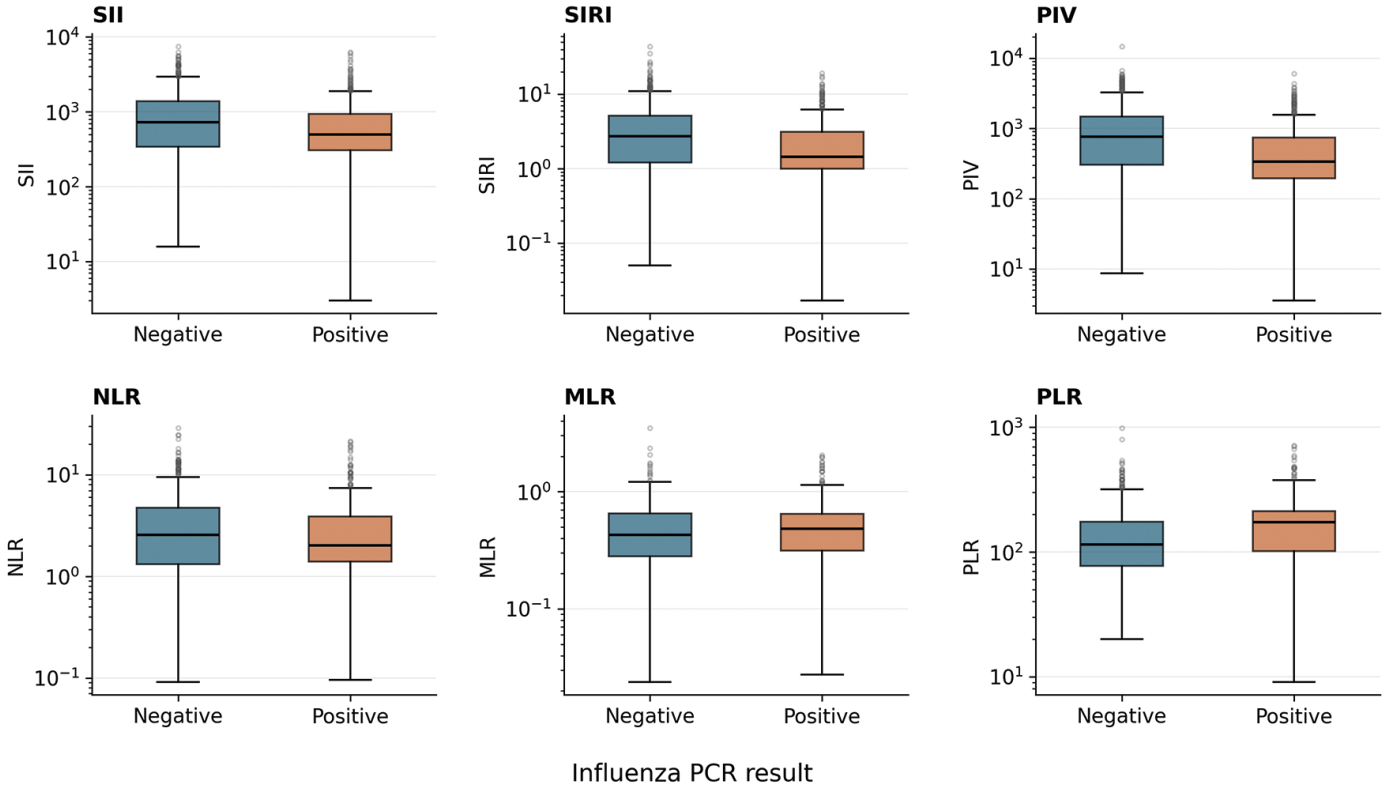


Figure 2. Distributions of blood count derived indices by influenza PCR status. The six panels show SII, SIRI, PIV, NLR, MLR, and PLR for all 903 children. Boxes delimit the 25th and 75th percentiles and horizontal lines show medians. Whiskers reach the most extreme observations within 1.5 times the interquartile range on the original scale; other observations are plotted individually. Vertical axes are logarithmic. Blue denotes influenza negative and orange influenza positive children

SII: Systemic immune-inflammation index, SIRI: Systemic inflammation response index, PIV: Pan-immune-inflammation value, NLR: Neutrophil-to-lymphocyte ratio, MLR: Monocyte-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, PCR: Polymerase chain reaction

Table 2. Blood count derived indices by influenza PCR status

Index	Negative (n=537)	Positive (n=366)	Raw p	Holm p
NLR	2.55 (1.32-4.74) [0.09-28.94]	2.02 (1.40-3.87) [0.10-21.45]	0.135	0.135
SII	723.6 (339.6-1387.5) [15.8-7496.5]	500.8 (306.2-943.4) [3.0-6253.1]	<0.001	<0.001
SIRI	2.73 (1.21-5.16) [0.05-43.89]	1.44 (1.00-3.10) [0.02-19.02]	<0.001	<0.001
PIV	765.4 (302.8-1483.6) [8.6-14703.4]	336.3 (195.4-742.3) [3.6-5990.5]	<0.001	<0.001
MLR	0.43 (0.28-0.65) [0.02-3.50]	0.48 (0.31-0.65) [0.03-2.04]	0.023	0.045
PLR	115.1 (77.1-174.2) [20.1-990.3]	174.2 (102.0-213.0) [9.1-717.6]	<0.001	<0.001
Hb/RDW	1.02 (0.91-1.12) [0.48-1.53]	1.05 (0.98-1.13) [0.32-1.59]	0.001	0.003

Data are median (25th-75th percentiles), with [minimum-maximum]. Quartiles use linear interpolation. Two-sided Mann-Whitney U tests compare distributions. Holm adjustment covers all seven index comparisons. Composite values use numerical counts in $10^3/\mu L$; hemoglobin is in g/dL and RDW in percent.
PCR: Polymerase chain reaction, NLR: Neutrophil-to-lymphocyte ratio, MLR: Monocyte-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, SII: Systemic immune-inflammation index, SIRI: Systemic inflammation response index, PIV: Pan-immune-inflammation value, Hb/RDW: Hemoglobin divided by red cell distribution width

Our lower leukocyte and platelet values are compatible with the pediatric findings of de Silva et al. (6), whose PCR-referenced cohort reported a leukocyte AUC of 0.71. Zhu et al. (4) evaluated age-specific blood count markers against symptomatic influenza-negative and healthy comparators, while Hong et al. (5) studied age-stratified CBC findings using an influenza A antigen reference test. Their populations, markers, and reference standards differ from ours, so diagnostic performance cannot be compared directly. These studies also reinforce that a negative respiratory comparator should not be treated as a healthy control.

The composite results should be interpreted based on their arithmetic values. SII equals platelets×NLR, SIRI equals monocytes×NLR, and PIV equals platelets×monocytes×NLR. Lower

platelet and monocyte distributions can therefore accompany lower composite indices, even when NLR itself does not separate the groups. In our data, lymphocyte counts were also lower, and MLR and PLR were higher in positive children, illustrating that the overall cellular pattern cannot be reduced to uniformly lower inflammation. Group medians do not establish which component caused an individual index value, and these cross-sectional data cannot demonstrate a specific immunologic mechanism.

Influenza-related hematologic studies address different clinical questions. Han et al. (7) and Qi et al. (17) evaluated diagnostic differentiation in populations containing adult patients and different comparator groups, while Kazancioglu et al. (18) and Yeşildağ and Kocabas (9) compared influenza with COVID-19. Ng et al. (8) studied pneumonia in intensive care, and Topaz et al. (19) examined RDW and hospital outcomes in influenza. Comparisons of pediatric SII and SIRI have also been reported for respiratory syncytial virus and *Mycoplasma pneumoniae* pneumonia (10). Their results do not provide a universal threshold for children with upper respiratory tract symptoms.

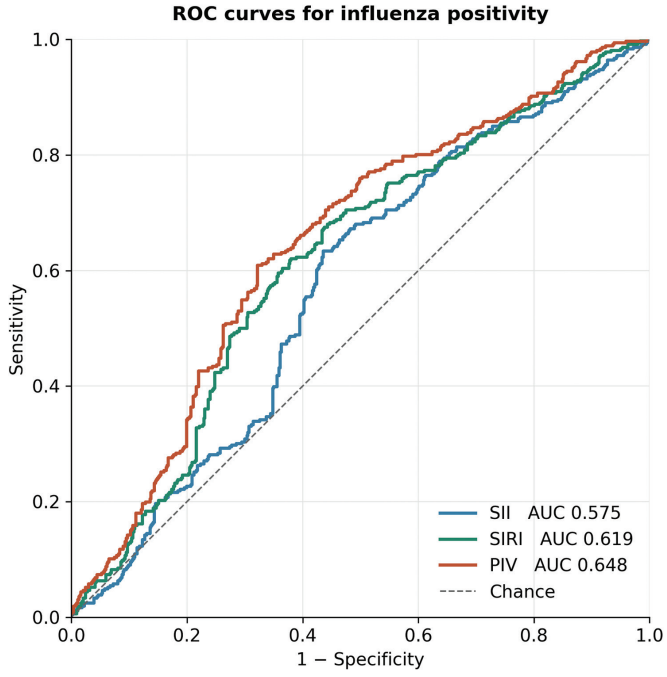


Figure 3. ROC curves for influenza positivity. All 903 children, including 366 positive children, are included. Lower index values indicate influenza positivity. Areas under the curve are 0.575 for SII, 0.619 for SIRI, and 0.648 for PIV. The dashed diagonal represents chance discrimination

ROC: Receiver operating characteristic, SII: Systemic immune-inflammation index, SIRI: Systemic inflammation response index, PIV: Pan-immune-inflammation value

Wu et al. (11) studied pulmonary complications among hospitalized influenza-positive children, using an aggregate index with the same neutrophil×monocyte×platelet/lymphocyte formula as PIV. That outcome differs from the influenza status among symptomatic ED attendees. Accordingly, a lower PIV in our positive group does not contradict an association between PIV and disease severity among infected patients, nor does it indicate that lower values correspond to milder disease. We did not measure disease severity or clinical outcomes.

The observed age imbalance materially affects the interpretation. No negative child was ≥10 years old; a full-cohort age coefficient would therefore extrapolate the comparator relationship to older positive children. Restricting adjusted analysis to common ages and allowing non-linear age and CRP effects provide a more bounded comparison. It cannot establish the association throughout childhood. PIV was selected during exploratory analysis, and conditional bootstrap validation does not remove the optimism associated with that selection. An improvement of approximately 0.03 in corrected AUC requires independent evaluation before any incremental diagnostic role can be asserted.

The exclusion procedure also shaped the comparison. PCT and cultures were obtained selectively, and the PCT threshold and positive-culture rule were operational criteria rather than constituting a comprehensive adjudication of bacterial infection. Influenza guidance emphasizes that PCT results require clinical interpretation and have limited specificity for bacterial coinfection (3). Exclusion of positive cultures, despite possible contamination, could remove children without true bacterial infections. Because the dataset contains data on only the included children rather than complete test denominators

Table 3. Exploratory regression within common ages

Predictor	df	Unadjusted OR (95% CI)	Global p	Adjusted OR (95% CI)	Global p
Age, spline	3	—	0.005	—	<0.001
CRP, $\log_2$ spline	3	—	<0.001	—	<0.001
PIV, per doubling	1	0.726 (0.663-0.796)	<0.001	0.746 (0.672-0.828)	<0.001

The post hoc sample contains 783 children aged 2.25-9.9167 years (537 negative, 246 positive). Adjusted analysis includes age, CRP, and PIV. Age and $\log_2$(CRP) use four-knot restricted cubic splines; neither has a single constant OR, indicated by —. PIV is $\log_2$-transformed, and the OR describes a doubling. OR confidence intervals use Wald variance; global p values use likelihood-ratio tests. The PIV Wald p value is also <0.001. Spline knots and coefficients are provided in the supplementary model specification.
OR: Odds ratio, CI: Confidence interval, df: Degrees of freedom, CRP: C-reactive protein, PIV: Pan-immune-inflammation value

Table 4. Model performance within common ages		
Measure	**Age + CRP**	**Age + CRP + PIV**
Parameters including intercept	7	8
AIC	897.05	866.66
BIC	929.69	903.96
Apparent AUC	0.708	0.740
Apparent Brier score	0.191	0.182
Corrected AUC	0.700	0.731
Corrected Brier score	0.194	0.185
Corrected calibration slope	0.949	0.953
Corrected calibration intercept	-0.033	-0.031

n=783 (537 negative, 246 positive), aged 2.25-9.9167 years. Models use the same age and $\log_2$(CRP) splines; the second adds $\log_2$(PIV). Corrected values are optimism-corrected estimates from 500 successful bootstrap resamples, re-estimating knots and refitting each model. Lower Brier scores indicate lower mean squared probability error. Ideal calibration slope and intercept are 1 and 0. Adding PIV: likelihood-ratio χ^2=32.39, df=1, p<0.001; apparent ΔAUC=0.032 and corrected ΔAUC=0.031. AIC and BIC are descriptive, not used for stepwise selection. Validation is conditional on the selected model and does not account for index selection.
AUC: Area under the receiver operating characteristic curve, AIC: Akaike information criterion, BIC: Bayesian information criterion, CRP: C-reactive protein, PIV: Pan-immune-inflammation value, df: Degrees of freedom

and clinical details for all screened patients, the direction and extent of this selection effect cannot be quantified.

Study Limitations

The retrospective single-center design and the requirement for CBC, CRP, influenza test results, and an adequate history limit generalizability. The exclusion of 243 children with incomplete laboratory or history data may also have introduced selection bias. Influenza-negative children had an unidentified mixture of etiologies in the analyzed dataset; results for other pathogens and alternative diagnoses were unavailable for characterization. Selective PCT/culture testing did not systematically exclude bacterial coinfection. The positive-culture or PCT >0.05 ng/mL exclusion rule could also have excluded children without true bacterial infection. The absence of negative children aged ≥10 years restricts adjusted inference to common ages. Index selection, sample-derived thresholds, model specifications, and sensitivity analyses were exploratory; internal validation was conditional on the chosen model, and no external validation was performed. Laboratory values were obtained from a single ED encounter; the analyzed dataset contained no serial measurements, precise sampling intervals, or symptom-duration information. Details of the assay platform and analyzer were not reported, limiting technical reproducibility. Treatment decisions, clinical severity, and patient outcomes were not evaluated.

Conclusion

Among children with upper respiratory tract symptoms in this selected ED cohort, influenza positivity was associated with lower SII, SIRI, and PIV, whereas NLR did not differ significantly between groups. These composite-index differences should be interpreted in the context of lower platelet and monocyte counts. Within overlapping age ranges, PIV remained inversely associated with positivity after flexible adjustment for age and CRP, and adding PIV improved the internally assessed model performance. However, the limited discriminatory power of individual indices and the sampling constraints preclude claims of validated clinical diagnostic utility. Prospective studies with clinically characterized comparator groups, broader age overlap, and external validation are needed to determine whether these indices add useful information to routine clinical assessment.

Ethics

Ethics Committee Approval: The study was approved by the Başkent University institutional review board (protocol no: KA25/476, date: 13.01.2026).

Informed Consent: The requirement for informed consent was waived by the Institutional Review Board due to the retrospective and anonymized nature of the study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: M.K., K.U.K., S.A.K., O.Ö., Concept: M.K., Design: M.K., Data Collection or Processing: M.K., K.U.K., S.A.K., O.Ö., Analysis or Interpretation: M.K., Literature Search: M.K., Writing: M.K.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The study was supported by the Başkent University Research Fund (project no: KA25/476).

Supplementary Tables 1, 2: https://d2v96fxpocvxx.cloudfront.net/a5223c9c-50f0-490d-b293-6a74c2af8d3b/content-images/10db83e8-720c-4378-8f39-10ec8362cea0.pdf

References

1. World Health Organization. Global influenza strategy 2019-2030. Geneva: World Health Organization; 2019. Available from: https://www.who.int/publications/i/item/9789241515320
2. Shang M, Blanton L, Brammer L, Olsen SJ, Fry AM. Influenza-associated pediatric deaths in the United States, 2010-2016. Pediatrics. 2018;141:e20172918.
3. Uyeki TM, Bernstein HH, Bradley JS, Englund JA, File TM, Fry AM, et al. Clinical Practice Guidelines by the Infectious Diseases Society of America: 2018 update on diagnosis, treatment, chemoprophylaxis, and institutional outbreak management of seasonal influenza. Clin Infect Dis. 2019;68:895-902.
4. Zhu R, Chen C, Wang Q, Zhang X, Lu C, Sun Y. Routine blood parameters are helpful for early identification of influenza infection in children. BMC Infect Dis. 2020;20:864.

5. Hong J, Lv J, Wu M, Shao J, Wu Q. The blood routine test holds screening values for influenza A in 2023: a retrospective study. Transl Pediatr. 2024;13:236-47.
6. de Silva MHAD, Ekanayake N, Darshana N. Validity of complete blood count in detecting influenza among clinically suspected children in Sri Lanka: a retrospective study from low resource setting country. BMC Infect Dis. 2026;26:816.
7. Han Q, Wen X, Wang L, Han X, Shen Y, Cao J, et al. Role of hematological parameters in the diagnosis of influenza virus infection in patients with respiratory tract infection symptoms. J Clin Lab Anal. 2020;34:e23191.
8. Ng WWS, Lam SM, Yan WW, Shum HP. NLR, MLR, PLR and RDW to predict outcome and differentiate between viral and bacterial pneumonia in the intensive care unit. Sci Rep. 2022;12:15974.
9. Yeşildağ M, Kocabas R. Influenza A (H1N1) and COVID-19 pneumonias: an evaluation in the light of new hematologic indices. J Inflamm Res. 2025;18:13103-16.
10. Zhou HH, Qian KL. Comparison of the systemic inflammatory response index and the systemic immune-inflammatory index in pediatric community-acquired pneumonia caused by respiratory syncytial virus and Mycoplasma pneumoniae. Front Pediatr. 2025;13:1694856.
11. Wu Y, Zhang J, Ke X, Pan J, Hu Y. Predictive value of the prognostic nutritional index and aggregate index of systemic inflammation for pulmonary complications in children with influenza. Transl Pediatr. 2026;15:343.
12. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. PLoS Med. 2007;4:e297.
13. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44:837-45.
14. Schuster NA, Rijnhart JJM, Twisk JWR, Heymans MW. Modeling non-linear relationships in epidemiological data: the application and interpretation of spline models. Front Epidemiol. 2022;2:975380.
15. Demler OV, Pencina MJ, D'Agostino RB Sr. Misuse of DeLong test to compare AUCs for nested models. Stat Med. 2012;31:2577-87.
16. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
17. Qi JF, Guo ML, Lin L, Fu S, Chen LL. An exploration of the value of NLR, PLR, LMR, and WBC × CRP for the diagnosis and treatment of influenza B in adults. Medicine (Baltimore). 2024;103:e37046.
18. Kazancioglu S, Bastug A, Ozbay BO, Kemirtlek N, Bodur H. The role of haematological parameters in patients with COVID-19 and influenza virus infection. Epidemiol Infect. 2020;148:e272.
19. Topaz G, Kitay-Cohen Y, Peled L, Gharra W, Kaminer K, Eitan M, et al. The association between red cell distribution width and poor outcomes in hospitalized patients with influenza. J Crit Care. 2017;41:166-9.