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: 24
Issue: 3

www.eajem.com

September
2025

galenos yayınevi

Official Journal of Emergency Physicians Association of Turkey

EPAT Emergency Physicians Association of Turkey

WACEM

ATUDER Acil Tıp Uzmanları Derneği

23-26 April 2026
Royal Seginus Hotel, Lara / Antalya

13TH INTERCONTINENTAL EMERGENCY MEDICINE CONGRESS
&
13TH INTERNATIONAL CRITICAL CARE AND EMERGENCY MEDICINE CONGRESS

IN CONJUNCTION WITH

22TH NATIONAL EMERGENCY MEDICINE CONGRESS
&
2026 WACEM SUMMER LEADERSHIP SUMMIT

EPAT ATUDER 27 YEARS Silver Jubilee 1999-2026

SCIENTIFIC SECRETARIAT

EPAT Emergency Physicians Association of Turkey

EMERGENCY PHYSICIANS ASSOCIATION OF TURKEY
Yukarı Ayrancı Güleryüz Sk. No:26/19
06550 / Çankaya / ANKARA
P: +90 312 426 12 14 F: +90 312 426 12 44
info@atuder.org.tr www.atuder.org.tr

ETHICON event & congress

ETHICCON EVENT & CONGRESS
Küçükbakkalköy Mah. Dereboyu Cad. No:3/A Brandium AVM R5 Blok D:22 K:3 Ataşehir / İstanbul
www.ethiccon.com

ORGANIZATION SECRETARIAT
NATIONAL CONGRESS
Mr. Yunus Özsümbül - Mail: yunus.ozsumbul@ethiccon.com - Phone: (+90) 536 255 40 40
INTERNATIONAL CONGRESS
Mrs. Gizem Özçelik - Mail: gizem.ozcelik@ethiccon.com - Phone: (+90) 538 845 38 98

EAJEM
Eurasian Journal of Emergency Medicine
Citation abbreviation: Eurasian J Emerg Med

OWNER AND RESPONSIBLE MANAGER ON BEHALF OF THE EMERGENCY PHYSICIANS ASSOCIATION OF TURKEY

Başar Cander
Department of Emergency Medicine, Bezmialem Vakıf University Faculty of Medicine, İstanbul, Türkiye
basarcander@yahoo.com
ORCID ID: 0000-0003-1035-1907

Editors in Chief

Editorial Board

Scientific Advisory Board

galenos yayınevi

Publisher Contact
Address: Molla Gürani Mah. Kaçamak Sk. No: 21/1 34093 İstanbul, Türkiye **Phone:** +90 (530) 177 30 97
E-mail: info@galenos.com.tr/yayin@galenos.com.tr **Web:** www.galenos.com.tr **Publisher Certificate Number:** 14521

Printing at: Son Sürat Daktilo Dijital Baskı San. Tic. Ltd. Şti.
Gayrettepe Mah. Yıldızposta Cad. Evren Sitesi A Blok No: 32 D: 1-3 34349 Beşiktaş, İstanbul, Türkiye **Phone:** +90 (212) 288 45 75
Printing Date: September 2025 **ISSN:** 2149-5807 **E-ISSN:** 2149-6048
International scientific journal published quarterly.

EAJEM
Eurasian Journal of Emergency Medicine
Citation abbreviation: Eurasian J Emerg Med

EAJEM
Eurasian Journal of Emergency Medicine
Citation abbreviation: Eurasian J Emerg Med

List of the reviewers who reviewed papers for Eurasian Journal of Emergency Medicine September 2025.

Alper Solakoğlu
Ayşe Ertekin
Behçet Varisli
Betül Çiğdem Yortanlı
David Adam
Demet Acar
Emine Emektar
Emine Özdemir Kaçer
Esra Ersöz Genç
Fatih Cemal Tekin
Gökhan Yılmaz
Hüseyin Mutlu
İlker Kaçer
Jacek Smeraka
Lukasz Chabowski
Mahmut Sami İnce
Melih Yüksel
Mert Özen
Mustafa Boğan
Osman Lütfi Demirci
Oya Güven
Sema Ayten
Serkan Günay
Şerife Özdinç
Turgut Dolanbay
Yeşim İşler

EAJEM
Eurasian Journal of Emergency Medicine
Citation abbreviation: Eurasian J Emerg Med

Please refer to the journal's webpage (**https://eajem.com/**) for "Aims and Scope", "Instructions to Authors" and "Ethical Policy".

The editorial and publication process of the Eurasian Journal of Emergency Medicine are shaped in accordance with the guidelines of ICMJE, WAME, CSE, COPE, EASE, and NISO. The journal is in conformity with the Principles of Transparency and Best Practice in Scholarly Publishing.

Eurasian Journal of Emergency Medicine is indexed in **Web of Science-Emerging Sources Citation Index, TUBITAK ULAKBIM TR Index, British Library, EBSCO, Gale, Embase, CABI, Directory of Research Journals Indexing, J-Gate, Türk Medline, Türkiye Atıf Dizini, DOAJ, Hinari, GOALI, ARDI, OARE, AGORA,** and **ProQuest**.

The journal is printed on an acid-free paper and published online.

Owner: Emergency Medicine Physicians Association of Türkiye

Responsible Manager(s): Başar Cander, Salvatore Di Somma, Sagar Galwankar

EAJEM
Eurasian Journal of Emergency Medicine
Citation abbreviation: Eurasian J Emerg Med

Contents

9 – 12 KASIM 2025
Arkin İskele Otel, KKTC

ÖĞLEN ÖĞREN tv

ATUDER
1999
Acil Tıp Uzmanları Derneği

12. ULUSLARARASI ACİL TIP VE İÇ HASTALIKLARI KONGRESİ

AKORT
AİLE HEKİMLİĞİ VE KORUYUCU HEKİMLİK TIP DERNEĞİ

6. ULUSAL AİLE HEKİMİ VE KORUYUCU HEKİMLİK TIP KONGRESİ

ETHICON
event & congress

T.C. SAĞLIK BAKANLIĞI
KONYA
İL SAĞLIK MÜDÜRLÜĞÜ

ATUDER
Acil Tıp Uzmanları Derneği

KTO KARATAY ÜNİVERSİTESİ

2nd INTERNATIONAL TRAUMA CONGRESS

KONYA CITY HOSPITAL

27-30th November 2025

Visit for info and registration:

acil42.com/**travmakongresi**

Send your abstracts via e-mail:

travmakongresi@gmail.com

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Editorial
DOI: 10.4274/eajem.galenos.2025.77785
Eurasian J Emerg Med. 2025;24(3): 150-1

Hematologic Inflammation Indices in Emergency CIN Risk Stratification in patients with Acute Coronary Syndrome-promise, Perspective, and Prudence

Sekhar Jyoti Sharma

All India Institute of Medical Sciences Faculty of Trauma and Emergency, Department of Trauma and Emergency, Guwahati, India

Keywords: ACS, contrast-induced nephropathy, hematologic indices, risk stratification

In the rapidly evolving and high-stakes environment of the emergency department (ED), clinicians often confront the dual imperatives of diagnostic efficiency and clinical prudence, particularly in patients presenting with acute coronary syndrome (ACS). The integration of contrast-enhanced imaging or percutaneous coronary intervention into the diagnostic and therapeutic algorithms for ACS has become ubiquitous. However, this advancement carries a known iatrogenic risk: contrast-induced nephropathy (CIN) (1). The need to balance life-saving interventions against preventable complications such as CIN underscores the necessity for rapid, cost-effective and accessible risk stratification tools.

CIN has been defined as the impairment of renal function gauged as either a 25% rise in serum creatinine from baseline or an increase of 0.5 mg/dL (44 µmol/L) in absolute serum creatinine value within 48-72 hours following intravenous contrast administration (2). The prevalence of diabetes and chronic kidney disease is rising by the day. Both of these are risk factors for acute kidney injury after cardiac catheterization and percutaneous coronary interventions. Based on current definitions the incidence of CIN ranges from 2% to 30%. Most cases are completely reversible within two to four weeks (3).

In this context, a study published in this issue, offers valuable insights by evaluating the prognostic utility of hematologic inflammation indicesnamely neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and lymphocyte-to-monocyte ratio (LMR) in predicting CIN among ACS patients in the emergency setting. This retrospective observational study of 814 patients identifies a modest but clinically relevant role for these biomarkers in early CIN risk stratification, with an emphasis on accessibility and rapidity.

What sets this study apart is not merely its focus on NLR, PLR, and LMR parameters that are increasingly recognized in the literature (4-7) but the real-world application in an emergency care context, where complex risk scoring systems often fall short due to time constraints or incomplete data. The authors aptly point out that most existing CIN prediction models (e.g., Mehran score) (8) are resource- and time-intensive, making them less practical in emergent settings. In contrast, inflammation indices derived from routine complete blood count offer a rapid, low-cost, and universally available alternative. The study's findings are notable. Patients who developed CIN (10.9% of the cohort) were significantly older and had higher NLR and PLR values, and lower LMR. Age >64 had the highest area under the ROC curve [area under the curve (AUC): 0.697], followed by NLR >5.2 (AUC: 0.615), PLR >137 (AUC: 0.590), and LMR <2.0 (AUC: 0.578). Importantly, NLR showed the highest specificity (84%), and LMR exhibited the highest sensitivity (86%) a compelling combination that underscores their potential complementary utility in ruling in or ruling out CIN risk.

Despite their statistical significance, the relatively modest AUC values of these markers underscore an important caveat: they are

Corresponding Author: Sekhar Jyoti Sharma, DA, DNB, All India Institute of Medical Sciences Faculty of Trauma and Emergency, Department of Trauma and Emergency, Guwahati, India
E-mail: sekharjyotis55@gmail.com **ORCID ID:** orcid.org/0009-0009-8553-1938

Cite this article as: Sharma SJ. Hematologic inflammation indices in emergency CIN risk stratification in patients with acute coronary syndrome-promise, perspective, and prudence. Eurasian J Emerg Med. 2025;24(3): 150-1.

Received: 01.09.2025
Accepted: 02.09.2025
Published: 10.09.2025

not definitive diagnostic tools, but rather adjunctive indicators. Their true strength may lie in their negative predictive value high across all indices which allows clinicians to confidently rule out CIN in low-risk patients and potentially reduce unnecessary testing, delays, or interventions.

Furthermore, these biomarkers may serve as a screening step prior to deploying more complex risk scores or initiating nephroprotective strategies, such as hydration protocols or minimizing contrast volume. The study thus aligns with the evolving concept of tiered risk assessment, where readily available data triage patients into different levels of monitoring or intervention intensity.

The study is grounded in a biologically plausible framework. Inflammation plays a central role in CIN pathogenesis, primarily through endothelial dysfunction, oxidative stress, and ischemic injury. NLR, PLR, and LMR are surrogate markers of this inflammatory milieu. Elevated NLR reflects neutrophilia and lymphopenia, both indicators of stress and systemic inflammation. A high PLR, meanwhile, suggests a pro-thrombotic, pro-inflammatory state, while a low LMR reflects monocytosis and suppressed adaptive immunity, both of which are implicated in tissue injury and impaired renal perfusion.

Like all retrospective single-center studies, this investigation is not without limitations. The absence of data on contrast type and volume, hydration status, and nephroprotective measures introduces potential confounding. Moreover, the use of only baseline values of the inflammatory markers overlooks the dynamic nature of inflammation, especially in acute settings like ACS. Nonetheless, the authors appropriately acknowledge these limitations, and their meticulous data screening and exclusion criteria lend credibility to their findings.

This study lays the groundwork for prospective, multicenter investigations that can validate these findings across diverse populations and healthcare systems. Future research should aim to:

- Integrate these hematologic indices into multivariable CIN prediction models.

- Examine serial measurements of NLR, PLR, and LMR for dynamic risk stratification.

- Evaluate the cost-effectiveness of using these markers in ED workflows.

Additionally, studies comparing these inflammatory indices with established scoring systems (e.g., Mehran, AKI risk index) could determine whether hybrid models offer improved performance.

Conclusion

This study provides timely and actionable insights into the use of inflammatory hematologic indices for CIN risk prediction in ACS patients undergoing contrast-enhanced procedures. While not definitive on their own, NLR, PLR, and LMR serve as practical, rapid-access tools that can assist emergency physicians in early decision-making. Their incorporation into clinical practice could enhance patient safety, optimize resource use, and support individualized care core pillars of modern emergency medicine.

As we continue to seek precision in patient care without compromising speed or simplicity, this study reminds us that sometimes, the answers lie not in more complexity, but in better use of what's already in front of us.

References

1. Yang Y, George KC, Luo R, Cheng Y, Shang W, Ge S, et al. Contrast-induced acute kidney injury and adverse clinical outcomes risk in acute coronary syndrome patients undergoing percutaneous coronary intervention: a meta-analysis. BMC Nephrol. 2018;19:374.
2. Moro AB, Strauch JGN, Groto AD, Toregeani JF. Creatinine level variation in patients subjected to contrast-enhanced tomography: a meta-analysis. J Vasc Bras. 2021;20:e20200161.
3. McCullough PA, Adam A, Becker CR, Davidson C, Lameire N, Stacul F, et al. Epidemiology and prognostic implications of contrast-induced nephropathy. Am J Cardiol. 2006;98:5K-13K. Epub 2006 Feb 10.
4. Yang Z, Qiao Y, Wang D, Yan G, Tang C. Association between inflammatory biomarkers and contrast-induced acute kidney injury in ACS patients undergoing percutaneous coronary intervention: a cross-sectional study. Angiology. 2024;75:831-40.
5. Karauzum I, Karauzum K, Acar B, Hanci K, Bildirici HIU, Kilic T, et al. Predictive value of lymphocyte-to-monocyte ratio in patients with contrast-induced nephropathy after percutaneous coronary intervention for acute coronary syndrome. J Transl Int Med. 2021;9:123-30.
6. Elcik D, Yarlioglu M, Oksuz F, Duran M, Inanc MT. Relationship between lymphocyte to monocyte ratio and contrast-induced nephropathy in acute coronary syndrome patients treated with percutaneous coronary intervention. J Med Res Rev. 2024;2:110-9.
7. Butt K, Souza J, Yuan C, Jayakumaran J, Nguyen M, Butt HI, et al. Correlation of the neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR) with contrast-induced nephropathy in patients with acute coronary syndrome undergoing percutaneous coronary interventions. Cureus. 2020;12:e11879.
8. Mehran R, Aymong ED, Nikolsky E, Lasic Z, Iakovou I, Fahy M, et al. A simple risk score for prediction of contrast-induced nephropathy after percutaneous coronary intervention: development and initial validation. J Am Coll Cardiol. 2004;44:1393-9.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.49358
Eurasian J Emerg Med. 2025;24(3): 152-62

Validity and Reliability of the Turkish Version of the Stanford Proxy Test for Delirium

Ebru Çakır Özden[1], Korhan Kollu[2], Dursun Elmas[2], Eda Ferahkaya[3], Sevim Bozkuş[4], Münevver Demirel[4], Jose Maldonado[5], Muhammet Cemal Kızılarslanoğlu[6]

[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 Intensive Care, Konya, Türkiye
[3]University of Health Sciences Türkiye, Konya City Hospital, Clinic of Pyschiatry, Konya, Türkiye
[4]University of Health Sciences Türkiye, Konya City Hospital, Clinic of Nursing, Konya, Türkiye
[5]Stanford University Faculty of Medicine, Department of Psychiatry and Behavioral Sciences, Consultation Liaison Psychiatry Service, Stanford, United States
[6]University of Health Sciences Türkiye, Konya City Hospital, Clinic of Internal Medicine, Division of Geriatrics, Konya, Türkiye

Abstract

Aim: The aim of this study was to perform a Turkish validation of the Stanford proxy test for delirium (S-PTDTV).

Materials and Methods: The original English version of the S-PTD was translated into Turkish using forward-backward translation methods. The Turkish version S-PTDTV was then administered by experienced nurses to elderly patients in an intensive care unit (ICU). The validation process involved assessing the sensitivity and specificity of the S-PTDTV by comparing its results with delirium diagnoses based on the diagnostic and statistical manual of mental disorders, fifth edition (DSM-5) and the confusion assessment method for the ICU (CAM-ICU). Reliability was assessed using internal consistency, intra-rater, and inter-rater reliability analyses.

Results: A total of 102 patients (50% female, mean age 74±9 years) participated in the study. When the cut-off score for the Turkish S-PTD was set at 7 points, the test showed a sensitivity of 96.6% and a specificity of 94.4% for the detection of delirium (area under the curve=0.985, $p<0.001$). High agreement was observed between S-PTD scores and both DSM-5 ($\kappa=0.885$, $p<0.001$) and CAM-ICU ($\kappa=0.932$, $p<0.001$). In addition, reliability analyses showed high consistency for both inter-rater [intraclass correlation coefficient (ICC=0.993, $p<0.001$)] and intra-rater (ICC=0.996, $p<0.001$) ratings. Internal consistency was also high, with a Cronbach's alpha of 0.914.

Conclusion: The results of this study indicate that the Turkish version of the S-PTD is a valid and reliable tool for the diagnosis of delirium in elderly ICU patients.

Keywords: A screening tool, delirium, intensive care unit, older patients

Introduction

Delirium and confusion are among the most common mental disorders in the elderly and medically ill patients. They are associated with many complex underlying medical conditions (1). Improving the recognition of delirium has the potential to reduce healthcare costs as well as patient morbidity and mortality, and minimise long-term adverse complications.

The diagnostic and statistical manual of mental disorders, fifth edition (DSM-5) criteria (2) are considered the gold standard for diagnosing delirium. However, many clinicians have difficulty with applying the DSM-5 criteria to diagnose delirium in clinical cases (1). Therefore, several diagnostic tools have been developed to help clinicians diagnose delirium, especially for non-psychiatric physicians (3). The most commonly used tools are the confusion evaluation method in the intensive care unit (CAM-ICU) and the delirium rating scale-revised-98 (DRS-R-98) (4,5).

Corresponding Author: Korhan Kollu MD, University of Health Sciences Türkiye, Konya City Hospital, Clinic of Internal Medicine, Division of Intensive Care, Konya, Türkiye
E-mail: korhankollu@gmail.com **ORCID ID:** orcid.org/0000-0002-0973-724X
Cite this article as: Çakır Özden E, Kollu K, Elmas D, Ferahkaya E, Bozkuş S, Demirel M, et al. Validity and reliability of the turkish version of the stanford proxy test for delirium. Eurasian J Emerg Med. 2025;24(3): 152-62.

Received: 29.01.2025
Accepted: 08.02.2025
Epub: 25.03.2025
Published: 10.09.2025

The Stanford proxy test for delirium (S-PTD) is a new delirium screening tool developed by Maldonado et al. (6), which is based on the recently published criteria of the DSM-5 and the 10th revision of the international classification of diseases and is designed specifically for use by nurses who follow patients throughout shifts (7-9). Compared with the data reported in the current literature, the S-PTD is more effective than other screening tools and is quicker and easier to use when administered under similar conditions. In addition, unlike other delirium screening tools, the use of the S-PTD is not hindered by the patient's unwillingness or inability to cooperate. This diagnostic tool is an important adjunct in the detection of delirium, improving patient care and allowing assessment by nurses, who spend the most time with patients and know them best (6).

The S-PTD consists of a total of 12 questions, which are scored as never (worth 0 points), sometimes (worth 1 point) and most of the time (worth 2 points). These parameters are attention, awareness/orientation, memory, communication, learning new information, reasoning and decision making, visuospatial difficulties, perception, disorganised thinking, behaviour and psychomotor activity, and sleep patterns. In addition, the time of development of all these changes, the fluctuation during the day and the age of the study participants were assessed. There is a 13-item, age, which is scored based on its actual numerical value (i.e, <55 y/o = "0", 56-70 y/o = "1", >70 y/o = "2"). To the best of our knowledge, the validity and reliability of the Turkish version of the S-PTD have not been investigated. Therefore, the aim of this study is to investigate the validity and reliability of the Turkish version of the S-PTD.

Materials and Methods

This study was conducted in the Internal Medicine and Nephrology Intensive Care Units at Konya City Hospital. Ethics Committee approval was obtained (decision number: 38/17, date: 17.12.2021).

The Study Protocol and Its Universe

Prior to commencing the study, Prof, Jose R. Maldonado was contacted by email to obtain copyright permission for the use of S-PTD, and the study commenced once permission had been obtained.

Stages of the Study

Forward-backward Translation Processes

First, the S-PTD was translated from its original language, English, into Turkish by professional native translators (forward translation). Then, the final version of the Turkish S-PTD was analysed by a team consisting of intensive care specialists, psychiatrists, internal medicine, and geriatrics specialists, and nurses, and compared with the original version in terms of meaning. The consensus version of the Turkish S-PTD was then reviewed by a linguist and the final version was produced.

The Turkish version of the S-PTD was then translated from Turkish back into English by another professional team, scientifically proficient in both Turkish and English, who had never read the original S-PTD before (back translation). The same team described above checked this version again. Finally, the backward translated form was compared with the original S-PTD for integrity of meaning. A linguist then reviewed the final version of the S-PTD to make any necessary adjustments. The final version of the S-PTD translated into Turkish can be found in Appendix 1.

Validity and Reliability Steps

Location and Population of the Study

This study was conducted at the Konya City Hospital. Patients aged 60 years and older who were treated in the internal medicine and nephrology intensive care units (ICUs) at the hospital with 45 beds were included in the study. Before enrolling, volunteers were given detailed information about the study. Subsequently, participants signed the informed consent form, by the patient or their relatives, and were enrolled in the study consecutively.

Sample Size and Statistical Power

ICUs are where delirium is most common. Some studies have reported that the incidence of delirium in elderly patients hospitalised in ICUs can be as high as 87% (4). Taking this into account, the sample size was calculated. The analysis was performed using OpenEpi version 3.01 (Andrew G. Dean and Kevin M. Sullivan, Atlanta, GA, USA). Our hospital has an ICU with 45 beds. It is predicted that the number of patients over the age of 60 who can be admitted to this ICU within three months will be around 200. Therefore, to achieve a 5% alpha error and a 95% confidence interval, with a design effect of 1, the minimum number of patients needed to reach power was determined to be 94.

Internal Consistency

The same nurse administered the Turkish version of the S-PTD to all patients included in the study. Cronbach's alpha coefficient was used to determine the internal consistency of the parameters of this test, with a value of 0.70 and above interpreted as indicating a strong level of consistency. As a result of the correlation analysis (Spearman's rank correlation test) between each parameter, the correlation coefficient of 0.81 and above was considered excellent, 0.61-0.80, was considered very good, 0.41-0.60, was considered adequate, correlations with a correlation coefficient between 0.21-0.40 were interpreted as

having an acceptable correlation, and values of 0.20 and below were accepted as an insufficient correlation.

Construct Validity

The diagnosis of delirium was assessed in three ways, and its validity was tested by examining the compatibility of these assessments. First, all patients in the study underwent a neuropsychiatric evaluation according to the DSM-5 criteria, which is considered the gold standard for the diagnosis of delirium. The DSM-5-based clinical assessment was performed by a team consisting of an internist, a psychiatrist, an intensive care physician, and a geriatrician. As a result of this assessment, all patients were categorized as having or not having delirium. In addition, patients diagnosed with delirium were further subdivided into hypoactive, hyperactive, or mixed delirium. This assessment was considered the gold standard. In addition, all patients were assessed by an internal medicine specialist using the CAM-ICU (10) as an objective test in this context. As previously described, patients were assessed with the S-PTD by the study nurse. The S-PTD and the gold standard assessment were administered within 60 minutes of each other. Members of the DSM-5-based diagnostic team and the study nurse were blinded to each other's diagnostic results and findings.

Patients with or without delirium were independently categorised according to the DSM-5 criteria, and according to the S-PTD. Agreement between the two assessments was analyzed using Kappa statistics. Kappa coefficient was considered: <0, no agreement; 0.0-0.20, insignificant agreement; 0.21-0.40, fair agreement; 0.41-0.60, moderate agreement; 0.61-0.80, significant agreement; 0.81-1.00, almost perfect agreement. In addition, receiver operating characteristic curve (ROC) analysis was performed to determine the best S-PTD estimate that could diagnose delirium using the gold standard method (DSM-5). An area under the curve (AUC) of 0.6 and above, and a p value of <0.05 were accepted as strong results.

Inter-rater Reliability

To test interrater reliability, two nurses were asked to perform the S-PTD on the same patients at different times (within 60 minutes). A total of 40 patients underwent this assessment. The S-PTD scores obtained from these assessments were tested by intraclass correlation coefficient (ICC) analysis. An ICC value of 0.70 or higher is regarded as indicating acceptable reliability.

Intra-rater Reliability

To test intra-practitioner reliability, a nurse was asked to perform the S-PTD on the same patients at two different times, with the second assessment occurring at least 30 minutes and no more than 60 minutes after the first assessment. A total of 40 patients underwent this assessment. ICC analysis was performed to test intra-rater reliability, and a value of 0.70 and above was considered acceptable.

Inclusion Criteria

All patients aged 60 years and older who were hospitalised in the ICU for at least 24 hours were consecutively included.

Exclusion Criteria

Patients under 60 years of age with a diagnosis of severe dementia, suspected delirium tremens, stupor/coma, intubation, acute cerebrovascular accident with intracranial haemorrhage, and unable to communicate were excluded from the study.

General Characteristics of The Patients

General demographic characteristics, education level, comorbidities, medications, reasons for ICU admission, date of ICU admission, and Acute Physiology and Chronic Health Evaluation-II (APACHE II) score were recorded. Chronic diseases were assessed using the Charlson comorbidity index (CCI). In addition, risk factors for delirium were assessed, including chronic diseases; presence of sepsis; acute vascular events; central nervous system pathologies; electrolyte imbalance; hypoxia; malnutrition risk (determined by the Nutritional risk screening-2002 score, which was routinely completed by the service nurses, with a score of three and above being accepted as a risk of malnutrition); dehydration (decided by physical examination); trauma; cancer; history of alcohol consumption; central venous catheter and urinary catheterisation.

Statistical Analysis

IBM SPSS version 21.0 package (Armonk, NY, USA) was used for statistical analysis. The Kolmogorov-Smirnov test, histogram, and coefficient of variation were used to test whether numerical variables were normally distributed. Normally distributed numerical variables were expressed as mean ± standard deviation, and non-normally distributed numerical variables were expressed as median (minimum-maximum). Categorical variables were presented as numbers and percentages.

When comparing numerical variables between independent groups, the Student's t-test was used for those with a normal distribution and the Mann-Whitney U test for those without a normal distribution. Chi-square or Fisher's exact tests were used to compare categorical data between independent groups. For reliability analyses, Cronbach's alpha was used for internal consistency, and ICC analysis was used to assess intra- and inter-practitioner reliabilities. Kappa and ROC analyses were used for validity. A p value of <0.05 was accepted as statistically significant.

Results

General Clinical Features and Delirium Status

The mean age of the patients was 74±9 years (50% female). The incidence of delirium was 29.4% and 27.5% according to DSM-5 and CAM-ICU criteria, respectively. The median values for the CCI, body mass index, and APACHE II score were 7, 25, and 21, respectively, while the number of medications used in the ICU was 5. The demographic data and general characteristics of the patients are summarised in Table 1.

When the S-PTDTV sub-parameters were evaluated in detail, it was found that all three categories were adequately scored in 11 out of 13 parameters. However, it was noted that item 9 (rating of disorganised thinking) was not adequately rated when the dimensions were explored. Therefore, the third column of this item was not scored after the assessments. On the other hand, it was noted that item-13, evaluation of age, was given 1 or 2 points. This is because we only conducted the study with older patients. The numbers and percentages of patients who responded to the S-PTDTV sub-parameters in detail are shown in Table 2. According to the DSM-5 criteria, the patients were divided into two groups with present or absent delirium and compared in terms of clinical characteristics (Table 3). As expected, the number of risk factors for delirium was higher in patients with delirium than in those without (p=0.045).

Validity and Reliability Analysis Results

ROC analyses showed that in our sample, the S-PTDTV had an overall sensitivity of 96.6% and specificity of 94.4% (p<0.001) for detecting delirium (compared with the DSM-5-based assessment/gold standard) when the cut-off was considered to be >6 (AUC=0.985). A difference was observed between hyperactive and hypoactive cases. In fact, these rates were 100% sensitivity and 97.22% specificity to detect hyperactive delirium (AUC=0.997, S-PTDTV cut-off >8, and p<0.001), while these rates had 87.5% sensitivity and 95.83% specificity in patients with hypoactive delirium (AUC=0.957, S-PTDTV cut-off >7, and p<0.001), compared to the gold standard DSM-5-based clinical assessment.

As the correlation coefficients of item-9 (disorganized thinking) and item-13 (age) were below 0.2 (as shown in the internal consistency analysis), the ROC analyses were repeated without the mentioned parameters. The cut-off for delirium were then found to be >5 when excluding item-13 only (AUC=0.984, p<0.001, 96.67% sensitivity and 95.83% specificity) or a different value when excluding both item-9 and 13 from the S-PTD (AUC=0.981, p<0.001, 96.7% sensitivity and 97.2% specificity). The results of the ROC curve analysis are shown in detail in Table 4.

Table 1. General clinical characteristics of the study population

Features	**p value**
Gender, female	51 (50.0)
Age, years	74±9
CCI	7 (2-13)
Number of drugs	5 (0-12)
BMI, kg/m^2	25 (12-47)
APACHE II score, median (min.-max.)	21 (8-48)
Comorbidities	
Hypertension	59 (57.8)
Diabetes mellitus	45 (44.1)
Coronary artery disease	22 (21.6)
Cerebrovascular events	9 (8.8)
Malignancy	22 (21.6)
Educational status	
İlliterate	52 (51.0)
Primary school graduate	42 (41.2)
Secondary school graduate	6 (5.9)
High school graduate	2 (2.0)
Reason for hospitalization	
Acute kidney injury	13 (12.7)
Sepsis	29 (28.4)
Respiratory failure	15 (14.7)
Hypervolemia	6 (5.9)
Pancreatitis	1 (1.0)
Gastrointestinal system bleeding	7 (6.9)
Other	31 (30.4)
Marital situation	
Married	69 (67.6)
Unmarried	2 (2.0)
Widowed	31 (30.4)
Smoking	
Unused	54 (52.9)
Smoker	25 (24.5)
Ex-smoker	23 (22.5)
Use of alcohol	
Unused	92 (90.2)
Active/social drinker	3 (2.9)
Ex-drinker	7 (6.9)
Delirium status according to DSM-5 criteria	
Present	30 (29.4)
Hypoactive	8 (7.8)
Hyperactive	16 (15.7)
Mixed type	6 (5.9)
Absent	72 (70.6)

Table 1. Continued	
Features	**p value**
Delirium status according to CAM-ICU criteria	
Present	28 (27.5)
Absent	74 (72.5)

Categorical variables were shown as numbers (n) and percentages (%). Normally distributed continuous parameters were presented as mean ± standard deviation while the skew distributed ones were as median (min.-max.). CCI: Charlson comorbidity index, BMI: Body mass index, CAM-ICU: Confusion assessment method for the ICU, DSM-5: Diagnostic and statistical manual of mental disorders, fifth edition, BMI: Body mass index, APACHE II: Acute physiology and chronic health evaluation-II, min.-max.: Minimum-maximum

Table 2. The numbers and rates of patients assessed for the S-PTD sub-parameters	
Parameters	**n (%)**
Item-1 (Attention)	
None	55 (53.9)
Sometimes	38 (37.3)
Most of the time	9 (8.8)
Item-2 (Awareness- orientation)	
None	66 (64.7)
Sometimes	26 (25.5)
Most of the time	10 (9.8)
Item-3 (Memory)	
None	64 (62.7)
Sometimes	25 (24.5)
Most of the time	13 (12.7)
Item-4 (Communication)	
None	76 (74.5)
Sometimes	19 (18.6)
Most of the time	7 (6.9)
Item-5 (Learning new information)	
None	67 (65.7)
Sometimes	27 (26.5)
Most of the time	8 (7.8)
Item-6 (Decision-making)	
None	75 (73.5)
Sometimes	21 (20.6)
Most of the time	6 (5.9)
Item-7 (Visuospatial)	
None	83 (81.4)
Sometimes	14 (13.7)
Most of the time	5 (4.9)
Item-8 (Perception)	
None	90 (88.2)
Sometimes	10 (9.8)
Most of the time	2 (2.0)
Item-9 (Disorganized thinking)	
None	97 (95.1)
Sometimes	5 (4.9)
Most of the time	-
Item-10 (Behavior or psychomotor activities)	
None	85 (83.3)
Sometimes	12 (11.8)
Most of the time	5 (4.9)
Item-11 (Sleep pattern)	
None	75 (73.5)
Sometimes	22 (21.6)
Most of the time	5 (4.9)
Item-12 (Fluctuation in severity)	
None	70 (68.6)
Sometimes	23 (22.5)
Most of the time	9 (8.8)
Item-13 (Age)	
≤55	-
56-70	37 (36.3)
>70	65 (63.7)

S-PTD: Stanford proxy test

In addition, inter- and intra-practitioner reliability analyses were evaluated using the ICC, which showed high reliability. The results are summarized in Table 5.

S-PTDTV scores were compared with DSM-5 and CAM-ICU scores and evaluated using Kappa concordance analyses. S-PTDTV cut-off were considered separately. The grading system proposed by Maldonado et al. (6), and according to the cut-off we found in our study (by removing age and disorganised items), was accepted. As a result of the Kappa analyses for each score, it was found that there was excellent agreement (Kappa values were in the range of 0.761-0.932) (Table 6). The internal consistency analysis showed a high Cronbach's alpha coefficient of 0.914 (Table 7). The correlation coefficients of item 9 (disorganised thinking) and item 13 (age) were below 0.2, indicating poor correlations.

Discussion

This study has shown that the Turkish version of the S-PTD is valid and reliable in detecting delirium when used in elderly patients admitted to ICU.

The incidence of delirium in ICUs is about 20 percent and the cumulative prevalence is almost 40 percent (11). More scales are currently being used to assess delirium in these patients (4).

However, simple methods that can be used by both clinicians and nurses are needed. As a result, the diagnosis of delirium is often missed by both doctors and nurses (12).

Our study evaluated the validity and reliability of the Turkish version of the S-PTD. It showed that, when the cut-off score was considered >6, the instrument had high sensitivity and specificity rates for the diagnosis of delirium (sensitivity 96%, specificity 94%, AUC=0.985). When items 9 and 13 were removed from the parameters, as the correlation coefficients were <0.2 in the internal consistency analysis, and the cut-off score was considered >5 in the diagnosis of delirium, the model was again found to have high sensitivity and specificity rates (sensitivity 96%, specificity 97%, and AUC=0.984).

In the previous study by Maldonado et al. (6), the sensitivity was 80.72% and the specificity was 90.37% when the cut-off score was >3 in the diagnosis of delirium; in the study by Alosaimi et al. (7), the sensitivity was 82.7% and the specificity was 95.3% when the cut-off score was >5. The sensitivity and specificity of the CAM-ICU test, another commonly used scale for the diagnosis of delirium, were 76-84% and 95%, respectively (13). Another commonly used tool for the diagnosis of delirium, the intensive care delirium screening checklist (ICDSC), has a reported sensitivity of 74-83% and specificity of 75-83%. Similarly, the DRS-R-98 has a reported sensitivity of 56-93% and specificity of 82-92%, while the Nursing delirium screening scale (Nu-DESC) has a reported sensitivity of 32-96% and specificity of 69-82% (5).

Table 3. Comparison of numerical and categorical data according to the delirium status

Parameters	Delirium present n=30	Delirium absent n=72	p value
Gender, female	12 (40.0)	39 (54.2)	0.192
Age, year	77±10	73±9	0.082
CCI	7 (2-11)	7 (2-13)	0.865
Number of drugs	4 (0-11)	5 (0-12)	0.515
BMI, kg/m^2	24 (12-37)	25 (12-47)	0.141
The number of delirium risk factors	5 (2-9)	4 (1-7)	**0.045**
APACHE II score	20 (11-41)	21 (8-48)	0.727
DM	13 (43.3)	32 (44.4)	0.918
HT	15 (50)	44 (61.1)	0.303
CAD	5 (16.7)	17 (23.6)	0.439
CVD	4 (13.3)	5 (6.9)	0.302
COPD	7 (23.3)	8 (11.1)	0.114
Malignancy	5 (16.7)	17 (23.6)	0.439

Categorical variables were shown as numbers (n) and percentages (%). Normally distributed continuous parameters were presented as mean ± standard deviation while the skew distributed ones were as median (min-max). DM: Diabetes mellitus, HT: Hypertension, CVD: Cerebrovascular disease, CAD: Coronary artery disease, BMI: Body mass index, CCI: Charlson comorbidity index, COPD: Chronic obstructive pulmonary disease, APACHE II: Acute physiology and chronic health evaluation-II

Table 4. The ROC curve analysis results

Parameters	AUC	Cut-off	p value	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)
S-PTD score (In recognizing delirium)	0.985	>6	<0.001	96.60	94.40	87.90	98.60
S-PTD score (In recognizing hyperactive delirium)	0.997	>8	<0.001	100.0	97.22	88.90	100.00
S-PTD score (In recognizing hypoactive delirium)	0.957	>7	<0.001	87.50	95.83	70.00	98.60
S-PTD total score (In recognizing delirium) (item-13 was not included in S-PTD)	0.984	>5	<0.001	96.67	95.83	90.60	98.60
S-PTD total score (In recognizing delirium) (item-9 and -13 were not included in S-PTD)	0.984	>5	<0.001	96.70	97.20	93.50	98.60

AUC: Area under the curve, NPV: Negative predictive value, PPV: Positive predictive value, S-PTD: Stanford proxy test, ROC: Receiver operating characteristic curve

Our current study found that using a cut-off score of >5 on the S-PTDTV scale (excluding parameters 9 and 13) gave the best results for diagnosing delirium. In addition, a higher sensitivity rate (96%) was observed than in the two previous studies (7). However, further studies may confirm this result, as the current study was conducted in a single centre with a small sample size and mostly older patients.

Although CAM-ICU, ICDSC, DRS-R-98, and other scales can be widely used to detect ICU delirium (6-10), the limited patient interaction due to ventilator dependence, especially in ICU patients with hypoactive delirium, may limit their use (14-17). Nevertheless, our findings from the current study, supported by the literature, suggest that the Turkish version of the S-PTD can be used quickly and safely in ICUs to assess delirium and its subtypes in hypoactive or hyperactive forms.

Table 5. Reliability analysis results

	ICC	p value
Inter-rater	0.993 (0.98-0.99)	<0.001
Intra-rater	0.996 (0.93-0.98)	<0.001

ICC: Intraclass correlation coefficient

The S-PTDTV can be applied by a nurse in as little as one minute (10). On the other hand, the Nu-DESC scale, one of the other scales used, has been reported to take approximately 1-2 minutes; the ICDSC scale approximately 3 minutes; and the CAM scale approximately 5 minutes (10,18,19). Therefore, the superiority of the S-PTDTV over other tests is due to its speed of use and because it can be administered by nurses; moreover, nursing practice skills and patient-nurse interaction may influence the results.

Age is known to be one of the most important risk factors for delirium (20), and the risk of delirium increases with age (21). In our study, the mean age of patients with delirium was higher than that of patients without delirium. However, this finding did not reach statistical significance (p=0.082) because only older patients and a relatively small sample size were included in the study. There is insufficient evidence on the effect of gender on the development of delirium (22). However, some studies have suggested that the risk of delirium may be higher in men (23). In our study, 60% of the delirium group was male, and there was no statistical difference compared with the non-delirium group.

Table 6. Validity analysis results

	The S-PTD's cut-offs in predicting delirium	Kappa	p value
DSM-5 vs. S-PTD	3 points and more	0.761	<0.001
DSM-5 vs. S-PTD	Higher than 6 points	0.885	<0.001
DSM-5 vs. S-PTD	When the age and disorganized substances were removed: higher than 5 points	0.930	<0.001
CAM-ICU vs. S-PTD	3 points and more	0.807	<0.001
CAM-ICU vs. S-PTD	Higher than 6 points	0.932	<0.001
CAM-ICU vs. S-PTD	When the age and disorganized substances were removed: higher than 5 points	0.931	<0.001

DSM-5: Diagnostic and statistical manual of mental disorders, fifth edition, CAM-ICU: Confusion assessment method for the ICU, S-PTD: Stanford proxy test

Table 7. Internal consistency analysis results

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
Item-1	-												
Item-2	0.650	-											
Item-3	0.657	0.829	-										
Item-4	0.653	0.570	0.659	-									
Item-5	0.699	0.619	0.708	0.781	-								
Item-6	0.568	0.663	0.725	0.661	0.750	-							
Item-7	0.509	0.535	0.628	0.662	0.584	0.617	-						
Item-8	0.353	0.248	0.313	0.475	0.433	0.318	0.408	-					
Item-9	0.297	0.187	0.288	0.333	0.351	0.186	0.157	0.493	-				
Item-10	0.405	0.315	0.267	0.441	0.381	0.258	0.497	0.572	0.257	-			
Item-11	0.576	0.488	0.394	0.459	0.484	0.321	0.414	0.512	0.278	0.715	-		
Item-12	0.594	0.582	0.524	0.604	0.641	0.490	0.557	0.588	0.351	0.680	0.790	-	
Item-13	0.135	0.113	0.158	0.136	0.019	0.034	0.105	0.055	0.171	0.039	-0.051	0.028	-

When the S-PTDTV sub-parameters were assessed in detail in our study, it was found that all three categories were assessed and scored appropriately in 11 out of 13 parameters. However, in the question assessing item 9 (disorganised thinking), most patients scored 0 and 1. This situation made us think that item 9 might be difficult for nurses to understand in Turkish. Item-13 (age) was scored only 1 or 2 points because we conducted our study on patients over the age of 60. For these reasons, the correlation coefficients of item 9 and item 13 in the internal consistency analysis were <0.2.

Therefore, subtracting these two parameters from one another may be an alternative way to evaluate older patients. However, the ROC and Kappa analyses, both with and without the two mentioned parameters, show that the S-PTD can be used in both ways. In the case of 13 parameters, the sensitivity and specificity for the diagnosis of delirium were 96% (AUC=0.985), when the cut-off was above 6 points. When 9 and 13 were subtracted from the parameters and the cut-off was considered >5 points, the sensitivity was 96% and the specificity was 97% (AUC=0.984).

Study Limitations

The main limitations include the relatively small sample size, that the study was conducted in a single centre, and that it only included patients aged 60 years and older. Replication studies with a larger sample, conducted in several medical centres, might show more accurate results. However, as our findings are supported by strong statistical results in the current study, we believe that our study will demonstrate the potential of S-PTDTV and stimulate further interest in this line of work.

Conclusion

The S-PTDTV is an efficient and easy-to-use delirium screening tool that is not affected by the fluctuating clinical variability of delirium, especially in clinical settings where patient cooperation is limited, such as ICUs, and other similar settings. In addition, the short time required to administer the S-PTDTV is likely to encourage obtaining information from the nurse rather than direct patient involvement in scoring. This study clearly demonstrates that the Turkish S-PTD is valid and reliable in assessing delirium in elderly patients hospitalized in ICUs.

Ethics

Ethics Committee Approval: This study was conducted in the Internal Medicine and Nephrology Intensive Care Units at Konya City Hospital. Ethics Committee approval was obtained (decision number: 38/17, date: 17.12.2021).

Informed Consent: Participants signed the informed consent form, by the patient or their relatives, and were enrolled in the study consecutively.

Footnotes

Author Contributions

Surgical and Medical Practices: E.Ç.Ö., K.K., D.E., E.F., S.B., M.D., J.M, M.C.K., Concept: E.Ç.Ö., K.K, D.E, E.F., S.B., M.D., J.M., M.C.K., Design: E.Ç.Ö., K.K., D.E., S.B., J.M., M.C.K, Data Collection or Processing: E.Ç.Ö., K.K., D.E., E.F., S.B., M.D., J.M., M.C.K., Analysis or Interpretation: E.Ç.Ö., K.K., D.E, E.F., S.B., M.D., J.M., M.C.K., Literature Search: E.Ç.Ö., K.K., D.E., S.B., J.M., M.C.K., Writing: E.Ç.Ö., K.K., D.E., S.B., J.M., M.C.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. Adamis D, Meagher D, Murray O, O'Neill D, O'Mahony E, Mulligan O, et al. Evaluating attention in delirium: a comparison of bedside tests of attention. Geriatr Gerontol Int. 2016;16:1028-35.
2. American Psychiatric Association D, Association AP. Diagnostic and statistical manual of mental disorders: DSM-5: American psychiatric association Washington DC. 2013.
3. Inouye SK. The dilemma of delirium: clinical and research controversies regarding diagnosis and evaluation of delirium in hospitalized elderly medical patients. Am J Med. 1994;97:278-88.
4. Cavallazzi R, Saad M, Marik PE. Delirium in the ICU: an overview. Ann Intensive Care. 2012;2:49.
5. van Velthuijsen EL, Zwakhalen SM, Warnier RM, Mulder WJ, Verhey FR, Kempen GI. Psychometric properties and feasibility of instruments for the detection of delirium in older hospitalized patients: a systematic review. Int J Geriatr Psychiatry. 2016;31:974-89.
6. Maldonado JR, Sher YI, Benitez-Lopez MA, Savant V, Garcia R, Ament A, et al. A study of the psychometric properties of the "stanford proxy test for delirium" (S-PTD): a new screening tool for the detection of delirium. Psychosomatics. 2020;61:116-26.
7. Alosaimi FD, Alghamdi A, Alsuhaibani R, Alhammad G, Albatili A, Albatly L, et al. Validation of the Stanford proxy test for delirium (S-PTD) among critical and noncritical patients. J Psychosom Res. 2018;114:8-14. Epub: 30.08.2018.
8. Chen TJ, Chung YW, Chang HR, Chen PY, Wu CR, Hsieh SH, et al. Diagnostic accuracy of the CAM-ICU and ICDSC in detecting intensive care unit delirium: a bivariate meta-analysis. Int J Nurs Stud. 2021;113:103782.
9. Plaschke K, von Haken R, Scholz M, Engelhardt R, Brobeil A, Martin E, et al. Comparison of the confusion assessment method for the intensive care unit (CAM-ICU) with the Intensive Care Delirium Screening Checklist (ICDSC) for delirium in critical care patients gives high agreement rate(s). Intensive Care Med. 2008;34:431-6.
10. Mitasova A, Kostalova M, Bednarik J, Michalcakova R, Kasparek T, Balabanova P, et al. Poststroke delirium incidence and outcomes: validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). Crit Care Med. 2012;40:484-90.

11. la Cour KN, Andersen-Ranberg NC, Weihe S, Poulsen LM, Mortensen CB, Kjer CKW, et al. Distribution of delirium motor subtypes in the intensive care unit: a systematic scoping review. Crit Care. 2022;26:53.
12. Wong CL, Holroyd-Leduc J, Simel DL, Straus SE. Does this patient have delirium?: value of bedside instruments. JAMA. 2010 Aug 18;304:779-86.
13. Ely EW, Inouye SK, Bernard GR, Gordon S, Francis J, May L, et al. Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). JAMA. 2001;286:2703-10.
14. Hargrave A, Bastiaens J, Bourgeois JA, Neuhaus J, Josephson SA, Chinn J, et al. Validation of a nurse-based delirium-screening tool for hospitalized patients. Psychosomatics. 2017;58:594-603.
15. Eren N. Delirium awareness and management in terms of personality traits of nurses working in intensive care units. Pamukkale Üniversitesi Sağlık Bilimleri Enstitüsü. 2019.
16. Lemiengre J, Nelis T, Joosten E, Braes T, Foreman M, Gastmans C, et al. Detection of delirium by bedside nurses using the confusion assessment method. J Am Geriatr Soc. 2006;54:685-9.
17. Detroyer E, Clement PM, Baeten N, Pennemans M, Decruyenaere M, Vandenberghe J, et al. Detection of delirium in palliative care unit patients: a prospective descriptive study of the Delirium Observation Screening Scale administered by bedside nurses. Palliat Med. 2014;28:79-86.
18. Radtke FM, Franck M, Schneider M, Luetz A, Seeling M, Heinz A, et al. Comparison of three scores to screen for delirium in the recovery room. Br J Anaesth. 2008;101:338-43.
19. Jung JH, Lim JH, Kim EJ, An HC, Kang MK, Lee J, et al. The experience of delirium care and clinical feasibility of the CAM-ICU in a Korean ICU. Clin Nurs Res. 2013;22:95-111.
20. Krewulak KD, Stelfox HT, Ely EW, Fiest KM. Risk factors and outcomes among delirium subtypes in adult ICUs: A systematic review. J Crit Care. 2020;56:257-64.
21. Pandharipande P, Shintani A, Peterson J, Pun BT, Wilkinson GR, Dittus RS, et al. Lorazepam is an independent risk factor for transitioning to delirium in intensive care unit patients. Anesthesiology. 2006;104:21-6.
22. Zaal IJ, Devlin JW, Peelen LM, Slooter AJ. A systematic review of risk factors for delirium in the ICU. Crit Care Med. 2015;43:40-7.
23. van Eijk MM, van Marum RJ, Klijn IA, de Wit N, Kesecioglu J, Slooter AJ. Comparison of delirium assessment tools in a mixed intensive care unit. Crit Care Med. 2009;37:1881-5.

Appendix 1. The Turkish Version of S-PTD (S-DTT Türkçe versiyonu)

Türkçe Stanford Deliryum Temsil Testi (S-DTT)

Maldonado JR ve arkadaşları. Psychosomatic 2020;61(2):116-26.

Stanford Üniversitesi Tıp Fakültesi, Psikiyatri Bölümü

Açıklama – Çalışma vardiyanız sırasında yapılan gözlemlere ve önceki 12 saat içinde diğer personel ve hastanın ailesi tarafından gözlemlenen veya bildirilen bilgilere dayanarak, aşağıda listelenen maddelerden herhangi birinin hastanız için geçerli olup olmadığını değerlendiriniz.	**Hiç**	**Bazen**	**Çoğu zaman**
1. Vardiyanız süresince hastanız <u>dikkat ile ilgili</u> zorluklar yaşadı mı? Örneğin; a) Soru sorduğunuzda veya yönlendirme yaptığınızda odaklanmada b) Konuşma esnasında dikkatinde kolayca dağılma c) Dikkat gerektiren görevlerde dikkatinde kolayca dağılma (örneğin; form doldurmak gibi)	0	1	2
2. Vardiyanız süresince hastanız <u>farkındalık/yönelim</u> ile ilgili zorluklar yaşadı mı? Örneğin; aşağıdakileri bilmeyle ilgili zorluklar: a) Nerede olduğunu b) Tıbbi durumunun ne olduğunu c) Neden burada olduğunu d) Tarihin ne olduğunu	0	1	2
3. Vardiyanız süresince hastanız <u>bellek ile ilgili</u> zorluklar yaşadı mı? Örneğin; a) Hastaneye neden başvurduğunu unutmak b) Ziyaretçi, öğün, prosedürler gibi günlük olayları unutmak c) Sağlık ekibi ve diğer personellerin kimliklerini/görevlerini unutmak gibi	0	1	2
4. Vardiyanız süresince hastanız sözlü veya yazılı <u>iletişim ile ilgili</u> zorluklar yaşadı mı? (Sadece konuşma değil) Örneğin; aşağıdakilerle ilgili zorluklar: a) Bir nesnenin ne olduğunu bilmek ancak nesnenin adını tam olarak hatırlayamamak b) Doğru kelimeleri anlamsız, saçma kelimelerle değiştirmek c) Sorulan sorulara mantıksız cevap vermek d) Anlaşılmayan şekilde veya mırıldanarak konuşmak	0	1	2
5. Vardiyanız süresince hastanız <u>yeni bilgi öğrenmede</u> zorluklar yaşadı mı? Örneğin; aşağıdakilerle ilgili zorluklar: a) Tıbbi durumu ile ilgili b) Fizyoterapi/Ergoterapi süresince yeni rehabilitasyon hareketlerini c) Yeni, hastane ekipmanlarını kullanmayı (mesela; yatak başı pisuvar, koltuk değneği, tekerlekli sandalye, aspiratör cihazı)	0	1	2
6. Vardiyanız süresince hastanız <u>mantıklı düşünme ve karar verme</u> konularında zorluk yaşadı mı? Örneğin; a) Sağlık ekibi veya ailesi ile bakım seçeneklerini tartışırken bilgileri mantıklı bir şekilde kullanmada b) Alternatifler önerildiği zaman tercih edilen seçeneği seçmede (mesela; yatağın konumlandırılması, jaluzilerin açık veya kapalı olması)	0	1	2
7. Vardiyanız süresince hastanız <u>görsel mekânsal (uzamsal)</u> zorluklar yaşadı mı? Örneğin; a) Yemek tepsisini getirip götürmede b) Bir şeyi tutarken kaybetme veya yerken, içerken, emerken ağzını bulamama gibi	0	1	2

Türkçe Stanford-Delirium Temsil Testi (S-DTT)
Maldonado ve arkadaşları, 2020

Açıklama – Çalışma vardiyanız süresince yaptığınız gözlemlere ve önceki 12 saat içinde diğer personel ve hastanın ailesi tarafından gözlemlenen veya rapor edilen bilgilere dayanarak "0" = "hiç", "1"= bazen, "2"= çoğu zaman" şeklinde derecelendiriniz.	**Hiç**	**Bazen**	**Çoğu zaman**
8. Vardiyanız süresince hastanız algılar ile ilgili zorluk yaşadı mı? Örneğin; a) İllüzyon (mesela; odadaki nesnelerin başka bir şey olduğuna inanmak veya duyduğu sesleri/konuştuğu dili yanlış yorumlamak) b) İşitsel ve/veya görsel Halüsinasyonlar (örneğin; derisini veya çarşafındaki "şeyleri" çekiştirme, hayali nesneleri tutma/işaret etme, odada olmayan insanlarla sohbet etme)	0	1	2
9. Vardiyanız süresince hastanız dezorganize (dağınık) düşünce sergiledi mi? Örneğin; a) Dağınık (dezorganize) konuşma veya konuyu dağıtma b) Gerçekle tutarsız olan sabit, yanlış inanışlar, mesela; • Paranoya (örneğin; sağlık ekibinin kendisini zehirlemeye çalıştığına dair inanışlar) • Grandiyöz (büyüklük) fikirler • Referans fikirler (örneğin; alakasız olayların hayatı için özel bir önemi olduğunu düşünür)	0	1	2
10. Vardiyanız süresince hastanız davranışlarında ve/veya psikomotor aktivitelerinde değişiklik sergiledi mi? Örneğin; a) Alışılmadık şekilde endişeli (ajite) ve aşırı uyarılmış (hiperalert) davranma (mesela; diken üstünde olma hali) b) Ruh halinde hızlı ve öngörülemeyen değişiklikler gösterme c) Alışılmadık şekilde yavaş hareket etme (düşünce veya hareketlerde), içe kapanma ve gözle görülür hareket eksikliği sergileyerek, üzgün veya depresif şekilde davranma	0	1	2
11. Vardiyanız süresince hastanız uyku düzeninde değişiklik gösterdi mi? Örneğin; a) Uykusuzluk yaşama b) Klinik olarak anlamlı olan ve günlük işlevlerini etkileyen gündüz aşırı uyku hali olması c) Gün içerisinde son derece canlı ve rahatsız edici rüyalar görme d) Rüyasındaki olayları gerçekte olmuş gibi anlatma	0	1	2
12. Yukarıda açıklanan değişiklikler, nispeten kısa bir süre içerisinde (saatlerden günlere) gelişmiştir ve hastanın başlangıçtaki dikkat ve farkındalık düzeyinden farklılık göstermektedir ve bu değişikliklerin şiddeti gün içerisinde dalgalanma eğilimindedir.	0	1	2
13.Yaş	≤55 yaş 0	56-70 yaş 1	>70 yaş 2
TOPLAM PUAN			

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.26879
Eurasian J Emerg Med. 2025;24(3): 163-70

Dominant vs. Non-dominant Hand in Pediatric Cardiopulmonary Resuscitation: A Randomized Crossover Simulation Trial

Burak Katipoğlu[1], Malgorzata Kietlinska[2], Michal Pruc[2], Alla Navolokina[3], Wojciech Wieczorek[4], Başar Cander[5], Lukasz Szarpak[6], Halla Kaminska[7], Monika Tomaszewska[2]

[1]Ufuk University Faculty of Medicine, Department of Emergency Medicine, Ankara, Türkiye
[2]LUXMED Group, Clinic of Research and Development, Warsaw, Poland
[3]International European University Faculty of Medicine, Department of Public Health, Kyiv, Ukraine
[4]Medical University of Warsaw Faculty of Medicine, Department of Emergency Medicine, Warsaw, Poland
[5]Bezmialem Vakıf University Faculty of Medicine, Department of Emergency Medicine, İstanbul, Türkiye
[6]The John Paul II Catholic University of Lublin, Faculty of Medicine, Institute of Medical Science, Lublin, Poland
[7]Medical University of Silesia Faculty of Medicine, Department of Children's Diabetology and Lifestyle Medicine, Katowice, Poland

Abstract

Aim: This study assessed the impact of hand dominance on the quality of pediatric chest compressions during simulated cardiopulmonary resuscitation (CPR).

Materials and Methods: A randomized crossover trial was conducted with 31 medical students trained in pediatric life support. Each participant performed chest compressions using both the dominant (DCC) and non-dominant hands (NDCC) on a high-fidelity pediatric simulator. Key CPR metrics, including compression depth, rate, hand placement accuracy, and rescuer fatigue, were analyzed.

Results: No statistically significant differences were found between DCC and NDCC in terms of compression depth ($p>0.05$), compression rate ($p>0.05$), rescuer fatigue ($p=0.864$), or perceived ease of compression ($p=0.612$). However, hand placement accuracy was significantly better with NDCC ($p=0.029$). Additionally, anthropometric factors, particularly body mass index (BMI) and height, positively correlated with compression depth and frequency, suggesting that individuals with higher BMI and height may achieve more effective compressions.

Conclusion: Hand dominance does not affect CPR quality, except for improved correct hand placement with NDCC. Personalized CPR training should consider rescuer characteristics to optimize performance. Further research is needed to refine pediatric resuscitation strategies.

Keywords: Pediatric cardiopulmonary resuscitation, hand dominance, chest compression quality, fatigue, medical simulation

Introduction

The efficacy of cardiopulmonary resuscitation (CPR) relies on the application of high-quality chest compressions, which are essential for the effective return of spontaneous circulation (ROSC) and post cardiac arrest neurological function, especially in children (1,2). The American Heart Association (AHA) and European Resuscitation Council guidelines recommend that, in adults, chest compressions should be 5-6 cm deep and delivered at a rate of 100-120 CPM, while in children, compressions should be one-third of the depth of the child's chest (3,4). Unfortunately, those standards can be quite difficult to implement in everyday clinical practice. Epidemiological studies indicate that the prevalence of out-of-hospital cardiac arrest in children varies from 8 to 20 per 100.000 individuals, with a survival rate to hospital discharge of 6 to 12% (5,6). Conversely, in-hospital cardiac arrest (IHCA) cases exhibit improved survival rates, ranging from 25% to 50%. However, the quality of CPR remains a critical determinant of long-term neurological outcomes (5,7).

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

Cite this article as: Katipoğlu B, Kietlinska M, Pruc M, Navolokina A, Wieczorek W, Cander B, et al. Dominant vs. non-dominant hand in pediatric cardiopulmonary resuscitation: a randomized crossover simulation trial. Eurasian J Emerg Med. 2025;24(3): 163-70.

Received: 14.02.2025
Accepted: 28.02.2025
Epub: 09.05.2025
Published: 10.09.2025

Insufficient quality of compressions results in hemodynamic implications. Previous studies have demonstrated that insufficiently deep compressions (less than 5 cm for adults) might reduce coronary perfusion pressure by 50%, hence diminishing the likelihood of ROSC (8). This scenario in children is further complicated by anatomical and physiological abnormalities. For instance, in pediatric intensive care unit (ICU) scenarios involving CPR, only 32.9% of instances adhered to the advised compression rate of 100-120 per minute. Several instances demonstrate that a chest compression rate of 80-100 compressions per minute (CPM) adversely correlates with survival, demonstrating an adjusted relative risk of 1.92 (9).

Interruption of chest compressions should be minimized to maximize efficiency. Compressions with pauses over 10 seconds reduce the survival probability by 3% for each five-second interruption (5). The quality of CPR diminishes owing to rescuer fatigue, which is prevalent even while adhering to established compression criteria within 2-3 minutes, for the inexperienced individual (8).

In children over one year of age, cardiac arrests mostly result from cardiac conditions such as congenital heart disease and cardiomyopathies (7). Furthermore, many comorbid variables significantly impact outcomes; patients with congenital cardiac disease constitute 21% of IHCA patients requiring extracorporeal cardiopulmonary resuscitation (ECPR), with a survival rate of 46%. Children lacking cardiac issues exhibited a 30% survival rate. In the past decade, the increased implementation of ECPR, especially in ICUs, has shown improved results; yet, it remains excessively dependent on the initial quality of CPR administered (6,7).

Despite advancements in CPR techniques, the impact of hand choice on compressions remains ambiguous. The one-hand chest compression technique is recommended for children and those with small thoracic dimensions; however, there are no guidelines on the preference for the dominant or non-dominant hand (NDCC) in its application. Certain data suggest that the dominant hand (DCC) may enhance the capacity to regulate compression force and depth, although it may also result in accelerated muscle fatigue (8). To enhance the efficacy of chest compressions, adherence to established parameters is essential; nevertheless, individual circumstances, including the rescuer's physique and the progression of tiredness, must also be taken into account. The integration of current research on one-handed CPR procedures with epidemiological data and pathophysiology considerations may lead to revised training protocols and enhanced worldwide survival rates.

The aim of this study was to evaluate and compare the quality of pediatric chest compressions performed using the DCC and NDCC hand in a simulated resuscitation setting.

Materials and Methods

This research was executed as a randomized crossover trial to assess and compare two pediatric chest compression methods: with DCC and the NDCC. The study was conducted in a regulated medical simulation environment, in compliance with international ethical standards for research involving human subjects. The Institutional Review Board of the Polish Society of Disaster Medicine provided ethical permission for the study (decision number: 44-2024-0312-IRB, date: 12.12.2024). All participants provided voluntary written informed consent before enrollment in the study.

The crossover design was deliberately employed to reduce inter-individual variability by having each participant serve as their own control, facilitating direct intra-individual comparisons among the three approaches. This analytical approach improves the accuracy of outcome evaluation and minimizes confounding variables related to individual skill disparities. We conducted the study in compliance with the CONSORT standards for reporting randomized crossover trials to uphold methodological integrity and enhance repeatability (10,11).

Participants

The study included medical students participating in pediatric life support (PLS) training conducted by AHA-accredited instructors. Inclusion criteria required prior completion of basic life support training (3). Exclusion criteria included recent upper limb injuries (within the past six months) and lack of prior exposure to pediatric CPR training.

Study Procedure

All participants were enrolled in a structured pediatric resuscitation training program, an essential part of their medical education. Before taking part in the study, they attended a standardized 60-minute theoretical session that covered the pathophysiology of pediatric cardiac arrest, CPR guidelines, and the technical details of two compression techniques under evaluation. These techniques were chest compressions performed with the DCC and those performed with the NDCC.

After completing the theoretical session, participants engaged in a practical training session under the guidance of an experienced instructor. This session involved hands-on demonstrations and supervised practice using an advanced pediatric simulator (MegaCode Kid, Laerdal Medical, Stavanger, Norway). To ensure proficiency and uniformity in technique, each participant was

given a 30-minute practice period before the experimental phase to become familiar with both compression methods. For the experimental phase, a different pediatric simulator (SimJunior®, Laerdal Medical, Stavanger, Norway) was utilized to maintain standardization and eliminate bias that could arise from performing CPR on the same simulator used during training. This approach helped prevent potential learning effects or unconscious adjustments to a familiar manikin, which might influence compression technique, depth, or consistency. The use of a separate, high-fidelity pediatric simulator for data collection ensured an objective evaluation of each technique's effectiveness, independent of familiarity with the equipment. To maintain measurement precision and reliability across all participants, the simulator was calibrated before each session. Additionally, to uphold consistency, the simulator was positioned on a firm surface in a well-lit environment.

Participants were randomly assigned to one of two groups, with one group beginning chest compressions using the DCC technique and the other starting with the NDCC technique (Figure 1). Each participant performed a continuous 2-minute compression cycle, followed by a 10-minute rest period to minimize fatigue and ensure sustained performance. The sequence continued in a crossover design until all participants had performed both techniques. This study design allowed each individual to act as their own control, facilitating a direct intra-individual comparison of compression efficacy, rescuer fatigue, and technique efficiency while reducing variability related to individual skill levels.

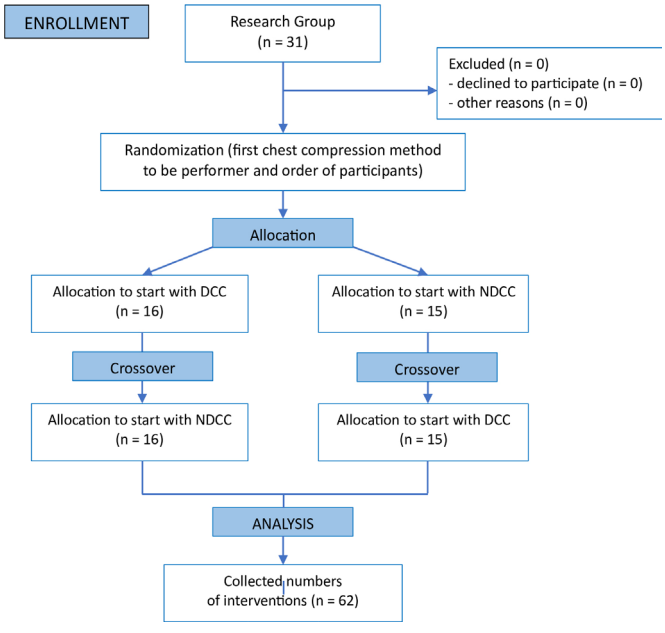


Figure 1. PRISMA flow chart
NDCC: Non-dominant hands, DCC: Dominant

Outcomes

The primary focus of this study was to evaluate the quality and effectiveness of chest compressions by analyzing key CPR performance metrics. These included the average compression rate, and the percentage of compressions performed within the recommended 100-120 CPM range, ensuring adherence to international guidelines. Compression depth was also examined, with an emphasis on the mean compression depth and the proportion of compressions reaching the recommended depth of ≥40 mm, as per the AHA guidelines (3). Additionally, chest recoil was assessed by measuring the percentage of compressions achieving full chest relaxation, which is crucial for optimizing coronary perfusion. Hand placement accuracy was recorded to determine the precision and consistency of chest compression techniques.

The subjective and physiological reactions of participants to different compression techniques were evaluated in more detail. Each compression method was rated on ease of performance, indicating how much effort was needed for each technique. Participants were asked to rate their discomfort and fatigue levels on the borg rating of perceived exertion as well as on the Numerical Rating scale from 0-10. A composite score that combined the objective measures of CPR performance and subjective measures of performance was calculated to assess both the quality of CPR and the effort needed from the rescuer.

All participants were required to perform CPR on the same pediatric simulator, which ensured that there was no variation in the participants' experiences. To reduce bias, all data were captured in real time via the feedback system of the pediatric simulator. This enabled the measurement of CPR quality to be independent of the participants' subjective insights.

Sample Size Determination and Statistical Power

The sample size was derived from power analysis guided by existing randomized crossover studies examining the efficacy of various chest compression techniques during pediatric resuscitation. The precision of estimates is increased by the crossover design, which reduces inter-personal variability, allowing every participant to act as their own control and making a smaller sample size possible.

Intervention research has shown a mean difference in compression depth of 3.0 mm [standard deviation (SD)= 2.5 mm] between different chest compression methods (12). Consequently, these estimates are selected as the outcome measure. A power calculation performed with G*Power 3.1 for a

within-subject ANOVA test showed that with alpha equal to 0.05, power set at 80 percent, and effect size set at (f) 0.25, at least 30 participants are required to detect observed differences in chest compression performance.

Due to the within-subject design, fewer participants are needed for this study than is typical with parallel-group designs. It is reasonable to assume that a sample of 30 subjects is sufficient to power statistical comparisons of compression depth, compression rate, and rescuer fatigue for different techniques.

Statistical Analysis

All statistical analyses were conducted using RStudio (Version 2024.12.0+467, R Foundation for Statistical Computing, Vienna, Austria). Descriptive statistics for continuous variables are presented as means and SD. The Shapiro-Wilk test was employed to assess the normality of data distributions, while Levene's test evaluated the homogeneity of variances. To compare participant performance between the two chest compression techniques (DCC vs. NDCC) within the framework of a randomized crossover design, a paired t-test was used. In instances where the assumption of normality was violated, the Wilcoxon signed-rank test served as a non-parametric alternative. The relationships between anthropometric parameters [gender, body weight, height, and body mass index (BMI)] and chest compression performance metrics were examined using Pearson's correlation coefficient (r) for variables exhibiting normal distributions and Spearman's rank correlation coefficient (p) for those not exhibiting normal distributions. The statistical significance of these correlations was determined, with a p value <0.05 considered indicative of significance. All statistical tests were two-tailed, and results are reported as Mean ± SD for both techniques, accompanied by the corresponding p values. Analyses were performed using the stats and rstatix packages in R. Data visualization, including boxplots, Bland-Altman plots, and correlation heatmaps, was executed using the ggplot2 package, facilitating a comprehensive comparison of the techniques.

Results

The study had 31 participants, whose average age was 21.19 years. The participants were found to have a mean body weight of 69.97 kg, a mean height of 171.90 cm, and a mean BMI of 23.59, ranging from 17.31 to 37.34, respectively. The subset of 19 female respondents had an average age of 21.42 body weight of 63.11 kg height of 166.11 cm BMI of 22.93. The remaining 12 participants, who constitute the rest of the sample, had characteristics that include being 21.03-year-old males with a body weight of 80.83 kg and a height of 181.08 cm, resulting in an average BMI of 24.64.

Results of Chest Compressions and Fatigue Outcomes

This included 31 participants who performed DCC and NDCC in a randomized crossover design. We analyzed descriptive statistics and the key performance parameters to compare DCC and NDCC techniques and compiled these data in Figures 2 and 3.

The ease of compression score averaged 4.29±2.47 for DCC and 4.13±2.01 for NDCC, exhibiting no statistically significant difference (p=0.612; Table 1). Similarly, there was no difference noted in the levels of fatigue (4.74±2.25 for DCC versus 4.81±2.32 for NDCC, p=0.864). The levels of hand pain also did not differ between the two techniques (4.32±2.23 versus 4.42±2.20, p=2.20).

On average, the performance score was better with NDCC, (68.58±30.26) compared to DCC (62.52±30.49). The difference, however, was not statistically significant (p=0.303). The average chest compression frequency did not differ statistically between the two techniques (115.81±16.05 versus 115.10±13.40, p=0.773).

The difference in the proportion of subjects with correct hand position was significantly higher for one technique (p=0.029), which shows that the particular method used influenced the hand positioning accuracy.

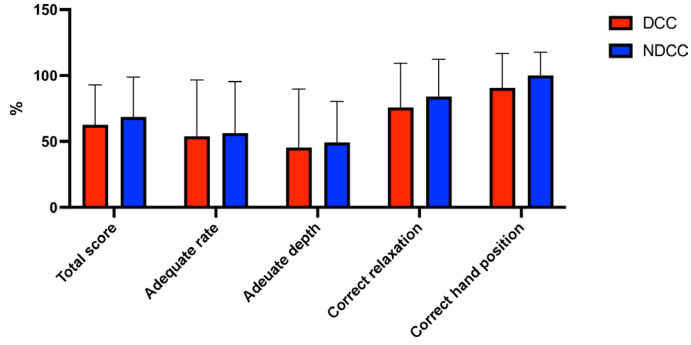


Figure 2. Chest compression quality among research groups
NDCC: Non-dominant hands, DCC: Dominant

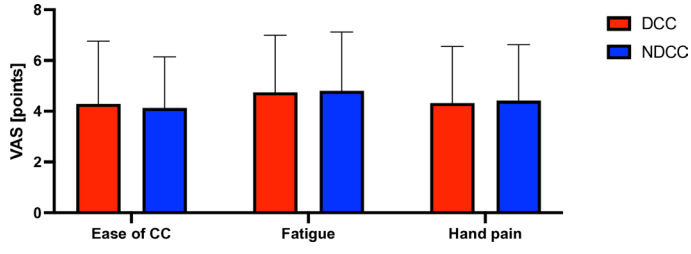


Figure 3. Effect of compression on fatigue and pain complaints
VAS: Visual Analog scale, NDCC: Non-dominant hands, DCC: Dominant, CC: Contralateral comparison

There were, however, no statistically significant differences in all other parameters except for correct hand position, which had a significant difference ($p>0.05$ for all other comparisons). This suggests that the perceived effort level, skill to perform compression, and the fatigue of the subjects were uniform across the methods.

Relationship Between Performance of Chest Compressions and Anthropometric Variables

Table 2 reveals the relationship between the anthropometric variables of gender, body weight, height, BMI, and chest compression performance. While weight and height may affect body compression depth and rate, other CPR components seem to be related to weight and height.

Table 1. Chest compression quality and fatigue outcomes among chest compression techniques

Parameter	Dominant hand	Non-dominant hand	p value
Total score	62.52 (30.49)	68.58 (30.26)	0.303
Mean rate	115.8 (16.04)	115.1 (13.4)	0.772
Adequate rate (100-120)	53.8 (42.77)	56.35 (39.14)	0.693
Mean depth	47.8 (7.04)	48.58 (11.23)	0.669
Adequate depth	45.32 (44.4)	49.26 (41.08)	0.604
Full relaxation	75.77 (33.58)	84.13 (28.35)	0.153
Correct hands position	90.61 (26.08)	100.0 (17.7)	0.029
Ease of chest compression	4.29 (2.46)	4.12 (2.01)	0.612
Fatigue level	4.74 (2.25)	4.81 (2.31)	0.864
Hands pain	4.32 (2.23)	4.41 (2.20)	0.781

Table 2. Correlation between anthropometric parameters and chest compression quality

Parameter	Gender differences	Effect of body weight	Impact of height	Body mass index
Dominant hand				
Total score	-0.065	0.135	0.086	0.104
Mean rate	-0.135	0.246	0.087	0.262
Adequate rate (100-120)	-0.014	-0.137	-0.151	-0.065
Mean depth	-0.289	0.387	0.396	0.212
Adequate depth	-0.303	0.329	0.415	0.118
Full relaxation	0.138	-0.333	-0.007	-0.358
Correct hands position	-0.084	0.287	-0.007	0.326
Ease of chest compression	-0.096	-0.011	-0.043	-0.016
Fatigue level	-0.032	-0.118	-0.070	-0.092
Hands pain	0.147	-0.268	-0.156	-0.226
Non-dominant hand				
Total score	-0.053	-0.004	-0.008	0.030
Mean rate	-0.039	0.236	-0.137	0.349
Adequate rate (100-120)	-0.099	-0.145	-0.048	-0.121
Mean depth	-0.287	0.349	0.285	0.249
Adequate depth	-0.155	0.205	0.214	0.118
Full relaxation	-0.103	-0.288	0.019	-0.362
Correct hands position	0.145	-0.083	-0.335	0.105
Ease of chest compression	-0.115	-0.085	0.080	-0.154
Fatigue level	-0.300	0.013	0.274	-0.151
Hands pain	-0.090	-0.049	0.166	-0.175

In terms of the total compression score, chest compression with the DCC technique showed weak positive relationships with body weight and BMI (r=0.136), suggesting that individuals with higher body mass tend to make more effective compressions. Furthermore, the average compression rate was positively related to BMI (r=0.262) and body weight (r=0.247), which indicates that participants with high BMI and body weight made a greater number of compressions. Compression depth was also moderately positively correlated with BMI, body weight (r=0.213), and height (r=0.396), indicating that deeper compressions were achieved by taller and heavier persons. Good compression depth was associated with BMI (r=0.118), body weight (r=0.329), and height (r=0.415). Additionally, there was a positive correlation between deep compressions and all other body composition tools.

The outcome presented in this section follows a similar pattern, with a difference in the strength of correlations. Total Skull Compression score showed almost non-existent correlation with BMI (r=0.030). Conversely, there was a moderate correlation between the mean compression and both BMI (r=0.349) and body weight (r=0.237). This suggests that those with a higher BMI and body weight performed a greater number of compressions, with BMI playing a greater role than it does with Technique A. Degree of compression showed positive correlation coefficients with BMI (r=0.250), body weight (r=0.350), and height (r=0.285), thus adding credence to the proposition that body mass and stature have an impact on compression effectiveness. In addition, good compression depth have positive r values with BMI (r=0.119), body weight (r=0.206), and height (r=0.215). As with technique A, the anthropometric measurements and the proper alignment, position, or frequency of hand compression did not correlate significantly, which suggests these aspects of performance are not greatly affected by body composition.

Discussion

This randomized crossover trial examines a significant gap in pediatric resuscitation research by assessing the influence of hand dominance on CPR quality in children, defined by the AHA as those aged one year to puberty. Our data contest traditional beliefs about the biomechanical advantage of the DCC in this age group, demonstrating that compression depth, rate, and rescuer fatigue are similar between DCC and NDCC compressions. The only exception-improved hand placement accuracy with NDCC-is especially pertinent in pediatric resuscitation, where anatomical limitations, that include smaller sternal landmarks and changing chest wall compliance, increase the likelihood of incorrect compression positioning. These findings correspond with recent research that supports method flexibility, while emphasizing the unique physiological and developmental factors specific to pediatric CPR.

The absence of notable variations in essential CPR parameters between DCC and NDCC contrasts with adult findings but aligns with previous pediatric studies (13). While prior research on baby resuscitation (<1 year) indicated comparable compression efficiency for dominant and non-dominant two-finger approaches, our findings broaden these insights to encompass older children. The findings indicate that hand preference does not inherently influence compression quality when appropriate biomechanics are upheld. This may indicate developmental adaptations: as children mature, the growing stiffness of the chest wall (attaining 40% of adult thoracic rigidity by age 8) requires greater compressive power, hence diminishing small variations in the effects of hand dominance (14). Nonetheless, the persistent dominance of NDCC in hand placement precision, especially vital in younger children with sternal lengths under 5 cm, indicates that cognitive-motor techniques may offset biomechanical constraints. Neuroergonomic studies suggest that the utilization of the non-dominant limb stimulates supplemental motor regions linked to intentional movement planning, perhaps improving accuracy when addressing delicate pediatric anatomy (13).

The enhanced hand placement accuracy reported with NDCC (p=0.029) has clinical significance due to the severe repercussions of improper positioning in pediatric patients. A recent registry investigation associated a lateral compression displacement of merely 1 cm in toddlers with a 22% rise in rib fracture rates and a 15% decrease in the likelihood of ROSC (15,16). Our data indicate that NDCC may provide a protective benefit, especially for rescuers with minimal pediatric experience. This corresponds with pediatric advanced life support standards that prioritize anatomical landmarks over compression force, although existing protocols do not address hand preference for children over one year of age. The divergence in placement accuracy and other performance metrics highlights the necessity for age-specific CPR quality standards-a notion that is gaining prominence in neonatal research but has not yet been systematically implemented in pediatric populations beyond infancy (13).

Anthropometric correlations indicated that rescuer BMI and height positively predicted compression depth, corroborating results from previous pediatric manikin research while raising distinct issues for application in older children. Although heavier rescuers attain deeper compressions, this “advantage” becomes a double-edged sword because of the significant variability in chest wall resilience among children aged 6 to 12 years (2.4-4.2 cm anteroposterior depth) (15). Universal depth targets (e.g., one-third chest depth) may elevate the danger of iatrogenic

damage if implemented without consideration of patient size and rescuer physique. Our data indicate that a 10 kg increase in rescuer weight is associated with 1.3 mm deeper compressions, a clinically significant difference considering that a 4 mm depth variation might affect cardiac perfusion pressure by 18% in school-age children. These findings support the implementation of weight-adjusted CPR algorithms similar to weight-based drug dosage methods, especially as juvenile patients near pubertal growth spurts that significantly affect chest biomechanics.

The contradictory relationship between objective performance and subjective weariness necessitates careful examination in pediatric resuscitation. Notwithstanding comparable compression measurements, participants indicated slightly elevated perceived exertion with NDCC an effect that was intensified in younger children during actual resuscitations. Prior research on pediatric CPR suggests that perceived weariness escalates as patient size diminishes, irrespective of actual exertion (17). This psychological-physiological disjunction may arise from the increased precision requirements of pediatric CPR as rescuers inadvertently apply compensatory muscular tension to prevent injury to perceived vulnerable patients. In pubertal patients nearing adult size, this impact may lessen, indicating that feelings of weariness are influenced by both the patient's size and the rescuer's experience (18).

These findings underscore the necessity for enhanced pediatric CPR instruction and protocols. For younger children (1-8 years), in which hand placement errors pose a significant risk, training could emphasize NDCC as a default strategy while retaining DCC as an alternative (19). In school-age youngsters, educators may advise on rescuer self-evaluation: "If landmarks are unclear, utilize the NDCC for accuracy; if depth is insufficient, return to the DCC for strength." For pubertal patients nearing adult physiology, adult CPR guidelines may become applicable earlier than formally specified, particularly in children exhibiting early pubertal development (20). Such age-stratified approaches align with the AHA's increasing recognition of developmental physiology in pediatric resuscitation but require further validation through multicenter trials (3).

Study Limitations

Despite the methodological rigor of this randomized crossover trial, several limitations must be acknowledged. First, the use of a simulated resuscitation environment, while necessary for standardization, may not fully replicate the complexities of real-world pediatric cardiac arrest scenarios. The controlled nature of the study, including the absence of factors such as emotional stress, variable patient presentation, and dynamic clinical environments, limits the generalizability of our findings to actual pediatric resuscitations. Second, the study cohort consisted exclusively of medical students who had undergone standardized PLS training. While this ensured a uniform baseline of CPR competence, it did not account for potential performance differences across a broader range of rescuers, including laypersons, paramedics, and pediatric specialists. Variability in clinical experience, strength, and technique among professional rescuers could influence CPR quality in ways not captured in this study. Third, the manikin used for data collection, while a high-fidelity simulator, may not perfectly replicate the biomechanical properties of a pediatric chest. Variations in thoracic compliance, rib elasticity, and sternal landmarks among real pediatric patients may impact compression accuracy and efficacy differently from those observed in this study. Future research incorporating cadaveric or live patient data is warranted to validate these findings. Fourth, although our crossover design minimized inter-individual variability, the 2-minute compression duration may not fully capture the dynamics of fatigue during prolonged resuscitations. In real-world settings, pediatric resuscitations often extend beyond this timeframe, potentially exacerbating rescuer fatigue and altering compression quality over time. Studies assessing longer compression durations and real-time fatigue accumulation are necessary. Fifth, the study did not account for potential handedness-related cognitive and motor adaptations over time. While our findings suggest no significant difference in CPR quality between dominant and NDCC use, prolonged training and experience may lead to compensatory strategies that could influence results. Longitudinal studies assessing the impact of repeated exposure and skill retention over time are needed. Finally, while our statistical power was sufficient for detecting differences in compression depth and accuracy, larger multicenter studies are required to confirm these findings across diverse populations and training backgrounds. Further investigation into how anthropometric factors influence CPR performance across different age groups and rescuer demographics could refine resuscitation guidelines.

Future investigations should emphasize three principal domains: 1) Biomechanical modeling of age-specific force-depth relationships throughout pediatric developmental stages; 2) Creation of adjustable CPR feedback systems that consider both patient age and rescuer anthropometrics; and 3) Longitudinal studies on technique retention in lay rescuers, noting that 67% of pediatric out-of-hospital cardiac arrests occur at home.

Conclusion

The study found no statistically significant differences in fatigue, pain, or ease of compression between the DCC and NDCC techniques, indicating that both methods require a similar level of effort and cause comparable discomfort. However, a

significant difference was observed in correct hand positioning, suggesting that one technique may offer an advantage in precision, which warrants further investigation. Additionally, anthropometric variables, particularly BMI, body weight, and height, showed positive correlations with chest compression depth and frequency. This suggests that individuals with greater body mass and stature are capable of performing deeper and more frequent compressions, potentially enhancing CPR effectiveness. However, factors such as hand alignment and compression frequency did not show significant associations with body composition, indicating that certain aspects of CPR performance are independent of physical attributes.

Ethics

Ethics Committee Approval: The Institutional Review Board of the Polish Society of Disaster Medicine provided ethical permission for the study (decision number: 44-2024-0312-IRB, date: 12.12.2024).

Informed Consent: All participants provided voluntary written informed consent before enrollment in the study.

Footnotes

Authorship Contributions

Concept: M.K, Design: M.K, L.S, Data Collection or Processing: M.K, W.W, L.S, Analysis or Interpretation: M.K, M.P, L.S, Literature Search: B.K, M.K, A.N, Writing: B.K, M.K, M.P, A.N, W.W, B.C, L.S, M.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. O'Halloran A, Nishisaki A. Understanding challenges to high-quality pediatric out-of-hospital cardiac arrest resuscitation performance. JAMA Netw Open. 2023;6:e2313931.
2. Kaminska H, Wieczorek W, Matusik P, Czyzewski L, Ladny JR, Smereka J, et al. Factors influencing high-quality chest compressions during cardiopulmonary resuscitation scenario, according to 2015 American Heart Association Guidelines. Kardiol Pol. 2018;76:642-7.
3. 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. Circulation. 2020;142(16_Suppl_2):469-523.
4. Van de Voorde P, Turner NM, Djakow J, de Lucas N, Martinez-Mejias A, Biarent D, et al. European nesuscitation council guidelines 2021: Paediatric life support. Resuscitation. 2021;161:327-87.
5. Lauridsen KG, Morgan RW, Berg RA, Niles DE, Kleinman ME, Zhang X, et al. Association between chest compression pause duration and survival after pediatric in-hospital cardiac arrest. Circulation. 2024;149:1493-500.
6. Farhat A, Ling RR, Jenks CL, Poon WH, Yang IX, Li X, et al. Outcomes of pediatric extracorporeal cardiopulmonary resuscitation: a systematic review and meta-analysis. Crit Care Med. 2021;49:682-92.
7. Sperotto F, Daverio M, Amigoni A, Gregori D, Dorste A, Kobayashi RL, et al. Extracorporeal cardiopulmonary resuscitation use among children with cardiac disease in the ICU: a meta-analysis and meta-regression of data through March 2024. Pediatr Crit Care Med. 2024;25:410-7.
8. Allan KS, Mammarella B, Visanji M, Moglica E, Sadeghlo N, O'Neil E, et al. Methods to teach schoolchildren how to perform and retain cardiopulmonary resuscitation (CPR) skills: a systematic review and meta-analysis. Resusc Plus. 2023;15:100439.
9. Sutton RM, Reeder RW, Landis W, Meert KL, Yates AR, Berger JT, et al. Chest compression rates and pediatric in-hospital cardiac arrest survival outcomes. Resuscitation. 2018;130:159-166.
10. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332.
11. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P; CONSORT NPT Group. CONSORT Statement for randomized trials of nonpharmacologic treatments: A 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. Ann Intern Med. 2017;167:40-7.
12. Tsou JY, Kao CL, Tu YF, Hong MY, Su FC, Chi CH. Biomechanical analysis of force distribution in one-handed and two-handed child chest compression-a randomized crossover observational study. BMC Emerg Med. 2022;22:13.
13. Gugelmin-Almeida D, Clark C, Rolfe U, Jones M, Williams J. Dominant versus non-dominant hand during simulated infant CPR using the two-finger technique: a randomised study. Resusc Plus. 2021;7:100141.
14. Niles DE, Nishisaki A, Sutton RM, Nysæther J, Eilevstjønn J, Leffelman J, et al. Comparison of relative and actual chest compression depths during cardiac arrest in children, adolescents, and young adults. Resuscitation. 2012;83:320-6.
15. Jiang C, Jiang S, Zhao Y, Xu B, Zhou XL. Dominant hand position improves the quality of external chest compression: a manikin study based on 2010 CPR guidelines. J Emerg Med. 2015;48:436-44.
16. You JS, Kim H, Park JS, Baek KM, Jang MS, Lee HS, et al. Relative effectiveness of dominant versus non-dominant hand position for rescuer's side of approach during chest compressions between right-handed and left-handed novice rescuers. Emerg Med J. 2015;32:184-8.
17. Badaki-Makun O, Nadel F, Donoghue A, McBride M, Niles D, Seacrist T, et al. Chest compression quality over time in pediatric resuscitations. Pediatrics. 2013;131:797-804.
18. Reynolds C, Cox J, Livingstone V, Dempsey EM. Rescuer exertion and fatigue using two-thumb vs. two-finger method during simulated neonatal cardiopulmonary resuscitation. Front Pediatr. 2020;8:133.
19. Cross J, Lam T, Arndell J, Quach J, Reed B, Thyer L, et al. Impact of hand dominance on effectiveness of chest compressions in a simulated setting: a randomised, crossover trial. Australas J Paramed. 2019;16:1-6.
20. Atkins DL, Berger S, Duff JP, Gonzales JC, Hunt EA, Joyner BL, et al. Part 11: pediatric basic life support and cardiopulmonary resuscitation quality: 2015 American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation. 2015;132(18 Suppl 2):519-25.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.32644
Eurasian J Emerg Med. 2025;24(3): 171-5

Evaluation of ChatGPTs Performance in Türkiye's First Emergency Medicine Sub-Specialization Exam

Hüseyin Mutlu, Kamil Kokulu, Ekrem Taha Sert, Muhammed Ali Topuz

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

Abstract

Aim: This study aims to evaluate ChatGPT's performance in Türkiye's Emergency Medicine Sub-Specialization Exam by assessing its success in answering both standalone and scenario-based questions through repeated testing.

Materials and Methods: This study utilized 60 multiple-choice questions from the Emergency Medicine Sub-Specialization Exam, comprising 30 standalone questions (50%) and 30 scenario-based questions (50%). Each question was presented to ChatGPT five times on different days, with all tests being conducted by researchers using the same computer. The latest version of ChatGPT, based on the GPT-4 architecture and extensively trained on medical texts and journals as of October 2023, was employed to ensure the highest available level of medical knowledge.

Results: ChatGPT achieved an overall accuracy rate of 85%, correctly answering 255 out of 300 questions across five trials. The accuracy rates for the five trials were 85% (51/60), 86.7% (52/60), 86.7% (52/60), 85% (51/60), and 81.7% (49/60), respectively, with no statistically significant difference between trials (p=0.94). ChatGPT demonstrated significantly higher accuracy in standalone questions compared to scenario-based questions [91.3% (137/150) vs. 78.7% (118/150), p=0.002]. Notably, ChatGPT exhibited consistent accuracy in interpreting visual data and correctly answering the two radiology-related questions across all five trials.

Conclusion: ChatGPT demonstrated high performance and consistency in Türkiye's first Emergency Medicine Sub-Specialization Exam, particularly excelling in standalone questions and radiological image interpretation. While the system is generally promising, its lower performance on scenario-based questions highlights the need for further development of clinical reasoning skills. These findings suggest potential applications of artificial intelligence systems in medical education and assessment, while emphasizing the necessity for improvements in clinical decision-making abilities.

Keywords: ChatGPT, artificial intelligence, emergency medicine, sub-specialty examination, medical education

Introduction

Recent advancements in artificial intelligence (AI) technology have introduced significant potential applications in the medical field. Specifically, Large Language models (LLMs), a subset of AI systems developed for natural language processing, have shown promise in medical knowledge evaluation and clinical decision-making (1). Among these, ChatGPT, developed by OpenAI and launched in November 2022, has attracted considerable attention for its performance in medical education (2).

Early studies evaluating ChatGPT's performance in medicine focused on the United States Medical Licensing Examination, where the system achieved success rates exceeding 60% (3). Similarly, studies conducted in Europe reported its successful performance in various specialty board exams (4). In the field of emergency medicine, ChatGPT has demonstrated that it can be used in successful triage of mass casualty events, and has shown promising results in Taiwan's Emergency Medicine Sub-Specialization Exam (5,6). While these findings highlight AI's potential in medical knowledge evaluation, they also underscore the need for further research on its role in clinical reasoning and decision-making processes.

Sub-Specialization Exams comprehensively assess both theoretical knowledge and clinical reasoning skills of the

Corresponding Author: Muhammed Ali Topuz MD, Aksaray University Faculty of Medicine, Department of Emergency Medicine, Aksaray, Türkiye
E-mail: alitopuzmd@gmail.com **ORCID ID:** orcid.org/0000-0003-2677-1260

Cite this article as: Mutlu H, Kokulu K, Sert ET, Topuz MA. Evaluation of ChatGPTs performance in Türkiye's first emergency medicine sub-specialization exam. Eurasian J Emerg Med. 2025;24(3): 171-5.

Received: 04.02.2025
Accepted: 14.03.2025
Epub: 05.05.2025
Published: 10.09.2025

candidates (7,8). Previous studies have reported ChatGPT's success in Sub-Specialization Exams for Medical Specializations (YDUS) and evaluations of sub-specialty trainees (7-9). In this study, we aim to evaluate ChatGPT's performance in Türkiye's first Emergency Medicine in YDUS (ED-YDUS). Specifically, we analyzed the system's performance on standalone and scenario-based questions, assessed its consistency across repeated tests, and examined its accuracy in radiological image interpretation.

Materials and Methods

In this observational study, we evaluated ChatGPT's performance on the ED-YDUS. In Türkiye, YDUS has been administered by the Student Selection and Placement Center (ÖSYM) across various sub-specialties since 2010. The first ED-YDUS was conducted on December 15, 2024 (10). The exam consists of 60 multiple-choice questions, each with five answer options, and is prepared in Turkish by ÖSYM. While designing the exam, ÖSYM refers to standard reference textbooks in emergency medicine, including Tintinalli's Emergency Medicine: Comprehensive Study Guide, 9th Edition, and Rosen's Emergency Medicine: Concepts and Clinical Application, 10th Edition. In this study, we used ChatGPT-4 Omni (ChatGPT-4o), considered to have the highest level of medical knowledge among its peers (11). ChatGPT-4o is an advanced LLM developed by OpenAI using the GPT-4 architecture (12), extensively trained on medical datasets, including texts and journals, up to September 2024 (13).

YDUS exam questions were obtained from the official website (https://ais.osym.gov.tr/bireyselgiriş/yandalsoruları) between December 15 and 25, 2024. The 60 multiple-choice questions were independently evaluated by two authors (Kamil Kokulu and Hüseyin Mutlu) and categorized into two groups: standalone questions and scenario-based questions. In cases where the two authors had differing decisions in categorization, a third author (Ekrem Taha Sert) reviewed the question, and a final decision was reached. Of the total questions, 30 (50%) were classified as standalone, while the remaining 30 (50%) were categorized as scenario-based.

Each question was presented to ChatGPT-4 one time on five separate days between December 25 and 31, 2024 by one of the authors (Muhammed Ali Topuz), using the same computer. For each question, five responses were generated. This methodology aligns with previous research where each question is presented three times to assess consistency and stability in responses generated by LLMs (10,14). All data, including the official answer key provided by ÖSYM, along with ChatGPT-4's responses, were systematically recorded in a Microsoft Excel 2023 document (Version 16.73, Microsoft Corporation, Redmond, WA).

Since this study did not involve human or animal subjects, ethical committee approval was not required.

Statistical Analysis

Statistical analysis was conducted using SPSS software (Version 26.0, SPSS Inc., Chicago, IL, USA). A $p<0.05$ was considered statistically significant. Categorical variables were presented as absolute numbers and percentages. Comparisons between categorical variables were performed using the chi-square test or Fisher's exact test as appropriate. Agreement between ChatGPT-4o's responses was evaluated using Cohen's kappa and Fleiss's kappa coefficients.

Results

ChatGPT correctly answered 255 out of 300 YDUS questions, achieving an overall accuracy rate of 85%. When presented with YDUS questions for the first time, ChatGPT demonstrated an accuracy rate of 85% (51/60). The accuracy rates for subsequent trials were 86.7% (52/60), 86.7% (52/60), 85%

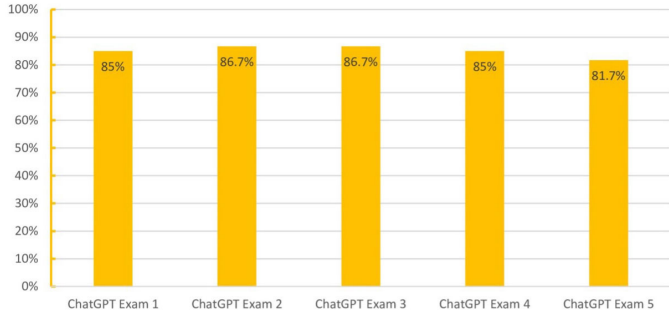


Figure 1. ChatGPTs Performances in Emergency Situation Medical YDUS exam

ED-YDUS: Emergency Medicine in Sub-Specialization Exams for Medical specializations

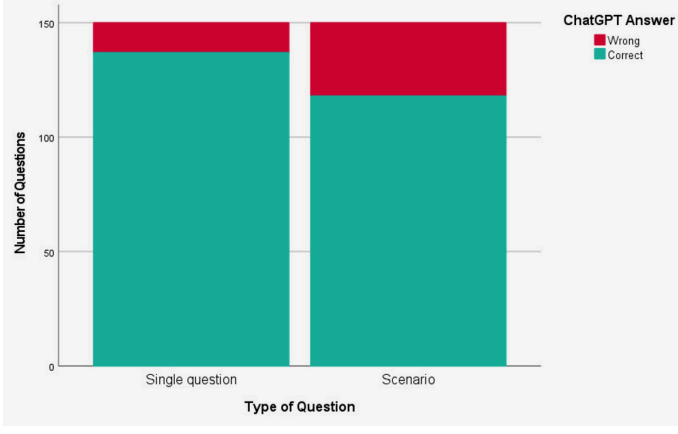


Figure 2. Accuracy of ChatGPT in ED-YDUS. ChatGPT 5 answers created for each question independently and scenario based

ED-YDUS: Emergency Medicine in Sub-Specialization Exams for Medical specializations

(51/60), and 81.7% (49/60), with no statistically significant difference in accuracy across trials (p=0.94) (Figure 1). ChatGPT demonstrated significantly better performance on standalone questions compared to scenario-based questions, with accuracy rates of 91.3% (137/150) and 78.7% (118/150), respectively (p=0.002, Pearson's chi-square test) (Figure 2). Additionally, ChatGPT consistently provided correct answers to both questions involving X-ray or computed tomography images and requiring radiological image analysis across all five trials (Figure 3).

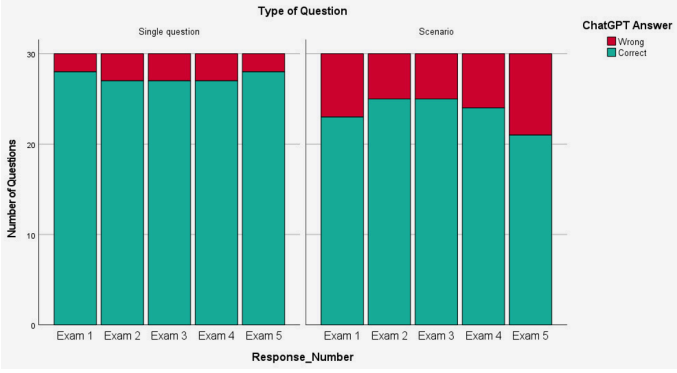


Figure 3. Five different exam ChatGPTs performances independent and scenario-based ED-YDUS questions

ED-YDUS: Emergency Medicine in Sub-Specialization Exams for Medical specializations

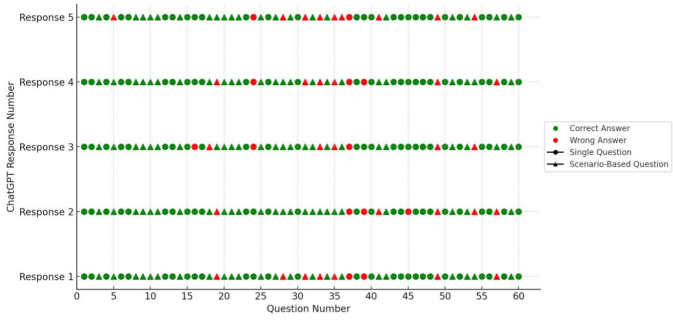


Figure 4. ChatGPT answer correctness by response number and question type

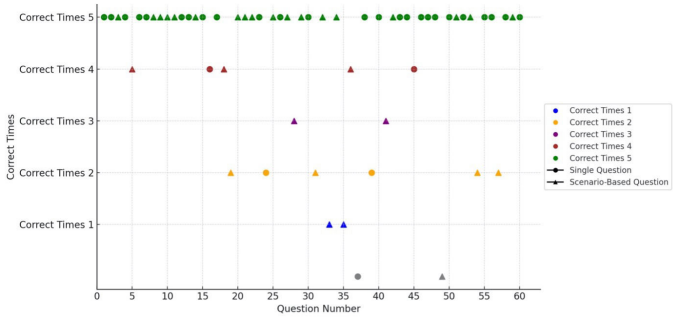


Figure 5. ChatGPT answer correctness by correct times and question type

Almost perfect agreement was found between the five answers (answers 1,2,3,4, and 5) given by ChatGPT for each question, when the answers correctly marked (or selected) by ChatGPT for each question (choice A,B,C,D, or E) were analyzed [Fleiss' kappa =0.83, 95% confidence interval (CI): 0.79-0.87, p<0.001] (Figure 4).

When the responses given by ChatGPT were categorised as "correct" and "incorrect", there was moderate agreement between the five responses given by ChatGPT for each question (Fleiss' kappa =0.50, 95% CI: 0.42-0.58, p<0.001) (Figure 5).

Of the exam questions: 12 Option A, 13 Option B, 11 Option C, 12 Option D, 12 Option E were correct answers. When the options marked by ChatGPT were analyzed, ChatGPT selected Option A 61 times, Option B 61 times, Option C 57 times, Option D 60 times, and Option E 61 times.

Discussion

The evaluation of medical knowledge and clinical reasoning skills of AI systems has become an increasingly important area of research in recent years (7,15). Our study is the first to evaluate the performance of ChatGPT in the first ED-YDUS in Türkiye. ChatGPT's overall success rate of 85% in five different exams and its consistent performance in repeated exams show the potential of AI systems in medical knowledge assessment. It is particularly noteworthy that all questions requiring radiological image analysis were answered correctly. This success parallels the development of the ability of AI systems to interpret visual medical data (7).

Ghanem et al. (9) reported that ChatGPT was less successful in scenario-based questions (55.56%) than in standalone questions (65.83%). Similarly, the study by Takagi et al. (16) found that ChatGPT performed less well in complex scenarios requiring clinical judgement. In our study, we also found that the performance of ChatGPT on standalone questions (91.3%) was statistically significantly higher than that on scenario-based questions (78.7%) (p=0.002). This finding suggests that AI systems are successful in assessing isolated medical knowledge, but have room for improvement in analysing complex clinical scenarios. This suggests that clinical decision-making processes should be further optimised in future versions of AI systems (1).

The fact that the distribution of ChatGPT's answers is balanced (A: 61, B: 61, C: 57, D: 60, E: 61) and that this distribution is similar to the distribution in the real answer key of the exam (A: 12, B: 13, C: 11, D: 12, E: 12) shows that the system does not respond randomly and makes a consistent evaluation. This finding supports the idea that AI systems have an objective and

systematic approach to the evaluation of medical information. A similar consistency was observed by Skalidis et al. (4) in their evaluation of the European Cardiology Board Examination, and it was emphasised that AI showed a systematic approach, choosing answers (2).

When analysing the performance of ChatGPT in repeated exams, success rates of 85%, 86.7%, 86.7%, 85%, and 81.7% were obtained from the first to the fifth exam, respectively. The study by Lee et al. (17) on Basic Life Support and Advanced Cardiovascular Life Support while Kokulu et al. (12) examined Paediatric Advanced Life Support. Both studies showed that ChatGPT-4 performed consistently in repeated assessments. This consistency suggests that AI systems can be used as a reliable tool in the assessment of medical information (18).

Ghanem et al. (9) reported that ChatGPT had little success with questions containing images, and Panthier and Gatinel (19) stated that questions containing images should be removed from scoring in their study conducted in the Ophthalmology Board Exam. In the study conducted by Toyama et al. (20) in the Radiology Board Exam, it was emphasised that AI systems still have areas requiring further development in image interpretation. In contrast to these studies, one of the interesting findings of our study was that ChatGPT correctly answered both questions requiring radiological image analysis, in all five repetitions. The success of ChatGPT in answering questions involving images shows the development of AI systems' skills in interpreting visual medical data. This is a remarkable development, especially given the importance of rapid and accurate interpretation of radiological images in emergency medicine.

Study Limitations

Our study shows that the latest GPT-4 model makes significant progress in addressing professional exam questions. However, there are some limitations to our study. First, as previous research has shown, the language in which questions are asked can have a significant impact on results. Since there may be a difference in meaning between the languages, when questions are asked in Turkish or translated into English, the accuracy of the answers may be affected. Secondly, if the sources of the questions (Tintinalli's Emergency Medicine: A Comprehensive Study Guide, 9th Edition, and Rosen's Emergency Medicine: Concepts and Clinical Practice, 10th Edition) were not taught to GPT-4, this may have influenced the accuracy of the answers. Third, since there were no Emergency Medicine Sub-Specialization study questions, they could not be uploaded to GPT-4. Finally, as analyses of the ED-YDUS results have not been published, the performance of GPT-4o could not be compared with actual exam results.

Conclusion

ChatGPT exhibited remarkable success in the first ED-YDUS exam in Türkiye with an overall success rate of 85% and an especially high performance in standalone questions with a success rate of 91.3%. Interestingly, it showed consistent performance in all five exams and correctly answered questions requiring radiological image analysis. The high success rate of ChatGPT in ED-YDUS, shows that AI systems can be used as a potential tool in medical education and knowledge assessment. However, the system's relatively low performance in scenario-based questions and clinical decision making suggests that AI cannot yet replace human clinical reasoning, but it can only be used as a supporting tool. Future research should focus on refining the clinical reasoning capabilities of AI systems and optimising their role in medical education.

Ethics

Ethics Committee Approval: Since the study did not involve human or animal subjects, ethics committee approval was not required.

Informed Consent: Since the study did not involve human or animal subjects, ethics committee approval was not required.

Footnotes

Authorship Contributions

Surgical and Medical Practices: H.M., K.K., Concept: H.M., K.K., M.A.T., Design: H.M., K.K., M.A.T., Data Collection or Processing: M.A.T., K.K., Analysis or Interpretation: H.M., K.K., E.T.S., Literature Search: H.M., K.K., E.T.S., M.A.T., Writing: H.M., K.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. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2:230-43.
2. Sharma S, Pajai S, Prasad R, Wanjari MB, Munjewar PK, Sharma R, et al. A critical review of ChatGPT as a potential substitute for diabetes educators. Cureus. 2023;15:e38380.
3. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ. 2023;9:e45312.
4. Skalidis I, Cagnina A, Luangphiphat W, Mahendiran T, Muller O, Abbe E, et al. ChatGPT takes on the European Exam in Core Cardiology: an artificial intelligence success story? Eur Heart J Digit Health. 2023;4:279-81.
5. Gan RK, Ogbodo JC, Wee YZ, Gan AZ, González PA. Performance of Google bard and ChatGPT in mass casualty incidents triage. Am J Emerg Med. 2024;75:72-8.

6. Liu CL, Ho CT, Wu TC. Custom GPTs enhancing performance and evidence compared with GPT-3.5, GPT-4, and GPT-4o? A study on the emergency medicine specialist examination. Healthcare (Basel). 2024;12:1726.
7. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28:924-33.
8. Noda R, Izaki Y, Kitano F, Komatsu J, Ichikawa D, Shibagaki Y. Performance of ChatGPT and Bard in self-assessment questions for nephrology board renewal. Clin Exp Nephrol. 2024;28:465-9.
9. Ghanem D, Nassar JE, El Bachour J, Hanna T. ChatGPT earns American Board Certification in hand surgery. Hand Surg Rehabil. 2024;43:101688.
10. https://www.osym.gov.tr/TR,5149/2010-ydus-sonbahar-donemi-ek-yerlestirme-tercihlerin-alinmasi-27122010.html
11. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt JN, Laleh NG, et al. The future landscape of large language models in medicine. Commun Med (Lond). 2023;3:141.
12. Kokulu K, Demirtaş MS, Sert ET, Mutlu H. ChatGPT and pediatric advanced life support: a performance evaluation. Resuscitation. 2024;205:110451.
13. He K, Mao R, Lin Q, Ruan Y, Lan X, Feng M, et al. A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics. Information Fusion. 2025;102963.
14. Zhu L, Mou W, Yang T, Chen R. ChatGPT can pass the AHA exams: open-ended questions outperform multiple-choice format. Resuscitation. 2023;188:109783.
15. Hoch CC, Wollenberg B, Lüers JC, Knoedler S, Knoedler L, Frank K, et al. ChatGPT's quiz skills in different otolaryngology subspecialties: an analysis of 2576 single-choice and multiple-choice board certification preparation questions. Eur Arch Otorhinolaryngol. 2023;280:4271-8.
16. Takagi S, Watari T, Erabi A, Sakaguchi K. Performance of GPT-3.5 and GPT-4 on the Japanese Medical Licensing Examination: comparison study. JMIR Med Educ. 2023;9:e48002.
17. Lee GU, Hong DY, Kim SY, Kim JW, Lee YH, Park SO, et al. Comparison of the problem-solving performance of ChatGPT-3.5, ChatGPT-4, Bing Chat, and Bard for the Korean emergency medicine board examination question bank. Medicine (Baltimore). 2024;103:e37325.
18. Delahanty RJ, Kaufman D, Jones SS. Development and evaluation of an automated machine learning algorithm for in-hospital mortality risk adjustment among critical care patients. Crit Care Med. 2018;46:e481-e488.
19. Panthier C, Gatinel D. Success of ChatGPT, an AI language model, in taking the French language version of the European Board of Ophthalmology examination: a novel approach to medical knowledge assessment. J Fr Ophtalmol. 2023;46:706-11.
20. Toyama Y, Harigai A, Abe M, Nagano M, Kawabata M, Seki Y, et al. Performance evaluation of ChatGPT, GPT-4, and Bard on the official board examination of the Japan Radiology Society. Jpn J Radiol. 2024;42:201-7.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.04468
Eurasian J Emerg Med. 2025;24(3): 176-82

The Role of the K/iCa Ratio in Predicting in Hospital Mortality in Patients with Upper Gastrointestinal Bleeding

Handan Özen Olcay[1], Emine Emektar[1], Şeref Kerem Çorbacıoğlu[1], Metin Uzman[2], Sedat Akkan[1], Yunsur Çevik[1]

[1]Ankara Atatürk Sanatorium Training and Research Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[2]Ankara Atatürk Sanatorium Training and Research Hospital, Clinic of Gastroenterology, Ankara, Türkiye

Abstract

Aim: Upper gastrointestinal bleeding (UGIB) is a common cause of emergency department presentation. Risk stratification of patients presenting to the emergency department with UGIB is important to predict active bleeding, need for blood transfusion, and mortality. In this study, we aimed to investigate the effect of potassium (K) and ionized calcium (iCa) on disease severity and mortality in UGIB patients.

Materials and Methods: This is a retrospective, single-center study. Patients aged 18 years and older who presented to the emergency department with UGIB, underwent blood gas analysis within the first half hour of presentation to the emergency department, were admitted to the hospital by the gastroenterology department, underwent endoscopy, and had complete data records were included in the study.

Results: The study included 699 patients. The in-hospital mortality rate was 7.7%. In the ROC analysis performed to determine the cut-off value of the K/iCa ratio for predicting mortality and active bleeding, the area under the curve values were 0.892 [95% confidence interval (CI): 0.845-0.939, $p<0.001$] and 0.832 (95% CI: 0.792-0.871, $p<0.001$), respectively. Patients were divided into 2 groups: those with a K/iCa ratio lower than 5.02 and those with a ratio higher than 5.02. This ratio has a high specificity value in terms of mortality prediction. Hemodynamic instability, active bleeding, need for erythrocyte suspension, length of hospital stay, and mortality were higher in the high K/iCa ratio group ($p<0.001$ for all values).

Conclusion: In patients presenting to the emergency department with suspected UGIB, assessment of the K/iCa ratio using blood gas analysis may help to predict the risk of active bleeding and in-hospital mortality as a rapid, inexpensive, and practical assessment method. Patients with a high K/iCa ratio may be candidates for early endoscopy and closer hemodynamic monitoring.

Keywords: Emergency department, upper gastrointestinal bleeding, in hospital mortality, K/iCa ratio

Introduction

Upper gastrointestinal bleeding (UGIB) is a common cause of emergency department (ED) visits. More than 70% of these bleedings are due to non-variceal etiologies, and gastroduodenal ulcer disease is one of the most common causes (1). Although pharmacologic and endoscopic treatments have been developed in recent years for the management of UGIB, it is still associated with significant morbidity and mortality. It can present with clinical symptoms ranging from massive bleeding leading to shock and death to occult bleeding leading to iron deficiency anemia from chronic blood loss (1,2). Risk stratification of patients presenting to the ED with UGIB is important to predict active bleeding, transfusion requirements, and mortality. Although various clinical scoring systems, biomarkers, and other risk factor assessments have been used for risk stratification, none of them have sufficient clinical reliability in routine clinical practice. Therefore, the search for simple, cheap and reliable prognostic markers continues.

We think that an important candidate for this research may be the ratio of potassium to ionized calcium (K/iCa ratio), the prognostic value of which has been the subject of recent studies, especially in critically ill and hemorrhagic patients. Calcium acts as a cofactor for many important functions, including the coagulation cascade, cardiac and smooth muscle contraction,

Corresponding Author: Handan Özen Olcay MD, Ankara Atatürk Sanatorium Training and Research Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
E-mail: hozen84@hotmail.com **ORCID ID:** orcid.org/0000-0002-1634-2684

Cite this article as: Özen Olcay H, Emektar E, Çorbacıoğlu ŞK, Uzman M, Akkan S, Çevik Y. The role of the K/iCa ratio in predicting in hospital mortality in patients with upper gastrointestinal bleeding. Eurasian J Emerg Med. 2025;24(3): 176-82.

Received: 01.02.2025
Accepted: 14.03.2025
Epub: 15.04.2025
Published: 10.09.2025

vasoconstriction, inotropy, and chronotropy; thus, calcium is essential for hemodynamic stability. Hypocalcemia can lead to coagulopathy in patients at high risk for bleeding. As in trauma patients, hypocalcemia is common in UGIB patients and may indicate the need for more blood transfusions and risk of death (3-5). iCa, the active form of serum calcium, can be measured directly and rapidly by blood gas analysis, which is available in many EDs and intensive care units.

Another electrolyte disturbance that may occur in hemorrhagic patients, including UGIB, is hyperkalemia. Hyperkalemia may be caused by mechanisms such as K release from red blood cells during degradation, tissue damage, decreased renal perfusion as a result of massive blood loss, and decreased urinary K excretion. The association of hyperkalemia with mortality in UGIB patients has been demonstrated in some studies (6). A recent study in traumatic hemorrhagic patients has shown that the K/iCa ratio has better prognostic value than the use of K or iCa alone (7).

We hypothesize that early K elevations, in addition to low iCa levels, may have a high predictive value for in-hospital mortality and active bleeding in UGIB patients. Therefore, in this study, we aimed to evaluate the prognostic value of K/iCa ratio on disease severity and mortality in UGIB patients.

Materials and Methods

Study Design

This study is a diagnostic-prognostic value study based on a retrospective cross-sectional analytical design and was conducted in the ED of Ankara Atatürk Sanatorium Training and Research Hospital, a 780-bed tertiary care center located in a large provincial center. The study was approved by the Local Ethics Committee Atatürk Sanatorium Training and Research Hospital (decision number: 20-KAEK15/2854, date: 27.12.2023). Our study was designed and reported in accordance with the Standards for Reporting of Diagnostic Accuracy Studies statement (8,9).

Data Collection

Data were collected from electronic medical records and patient charts. The chart review (clinical and demographic characteristics of the patients, including endoscopy results) was performed retrospectively by two different emergency physicians with at least 5 years of experience (9).

Study Population

We analyzed the consecutive database of all patients who presented to the ED with UGIB and underwent endoscopy performed by the gastroenterology department between January 2019 and January 2024. Patients aged 18 years and older, who presented to the ED with UGIB, underwent blood gas analysis within the first half hour of presentation to the ED, underwent endoscopy after admission to the gastroenterology department, and had complete data records were included in the study. Pregnant women were excluded from the study as well as patients who did not undergo endoscopy, those with lower GI bleeding, chronic renal failure, known parathyroid disease, known gastric cancer, indications for hospitalization other than UGIB, referred from another center, and missing data. In addition, the blood gas K level was compared with the biochemical K level. Patients with a biochemical hemolysis index greater than 1 were excluded.

Patients were classified in the hospital data system according to international statistical classification of diseases, tenth revision codes (K25, K25.4, K26, K26.4, K27, K27.4, K28, K28.4, K92.1). Demographic information, comorbidities (chronic hypertension, diabetes mellitus, coronary artery disease, liver disease), medications, admission vital signs, laboratory results, endoscopy results, blood and blood product indications, hospitalization, and mortality were recorded. Blood gas samples were analyzed using the Siemens RAPIDLab 1265 device within 5 minutes of collection. A systolic blood pressure of less than 90 mmHg, a heart rate of 100 beats/min or more, and/or a history of syncope with onset of symptoms were considered to be hemodynamically unstable. Gastritis, gastric/duodenal ulcer, esophageal varices, gastric cancer, Dieulafoy's lesion, Mallory-Weiss tear, angiodysplasia, and active bleeding were analyzed in the endoscopy results. Treatments applied during endoscopy, such as sclerotherapy, hemoclip, heater probe, and band ligation, were recorded. The K/iCa ratio of the patients was obtained from the results of blood gas analysis. Demographic and clinical data of patients were compared in relation to the K/iCa ratio and mortality status (in-hospital mortality).

Index and Reference Tests

The index test was determined as K/iCa ratio, and the primary reference test was determined as in-hospital mortality status of patients. In addition, active bleeding after endoscopy was determined as a secondary reference test. Active bleeding was defined as forest classification 1a, 1b, and active esophageal varices bleeding.

Statistical Analysis

All data collected throughout the study and recorded on the study form were analyzed using IBM SPSS 20.0 (Chicago, IL, USA). The Kolmogorov-Smirnov test was used to determine whether the distribution of discrete and continuous numerical variables was normal. Descriptive statistics were presented as median [interquartile range (IQR): 25-75] for discrete and continuous numerical variables, and as the number of cases (%) for categorical variables. Categorical variables were evaluated

by the chi-squared test and continuous variables by the Mann-Whitney U test. The ROC curve and area under curve (AUC) were calculated to determine the predictive and threshold values of K/iCa ratio for in-hospital mortality and active bleeding (10). The critical alpha value was considered 5% for all statistical analyses.

Results

During the study period, 1388 patients diagnosed with UGIB were evaluated. The study included 699 patients with complete data (Figure 1). 68% of patients were male and the median age was 63 years (IQR: 25-75, 47-74).

Endoscopic findings and endoscopic interventions are shown in Table 1. The most common endoscopic finding was gastritis (71.1%); gastric/duodenal ulcer was observed in 69.5% of patients. The percentage of active bleeding was found to be 18.7% and the in-hospital mortality rate was 7.7%. Among patients with gastric/duodenal ulcers (n=486), 2.1% were classified as Forrest 1a, 18.9% as Forrest 1b, 9.9% as Forrest 2a, 29.2% as Forrest 2b, and 23% as Forrest 2c.

The median K/iCa ratio was 3.85 (IQR: 25%-75%: 3.43-4.24) in the survivor group and 4.85 (IQR: 25%-75%: 4.44-5.26) in the non-survivor group, and the difference was statistically significant (p<0.001). Other clinical and demographic characteristics of the survivor and non-survivor patients are presented in Table 2. In addition, the median K/iCa ratio was 3.80 (IQR: 25%-75%: 3.38-4.15) in patients without active bleeding, and 4.6 was (IQR: 25%-75%: 4.16-5) in patients with active bleeding, and the difference was statistically significant (p<0.001) between the two groups.

In the ROC analysis performed to determine the cut-off value of the K/iCa ratio for predicting mortality and active bleeding, the AUC values were 0.892 [95% confidence interval (CI): 0.845-0.939, p<0.001] and 0.832 (95% CI: 0.792-0.871, p<0.001), respectively (Figure 2). The best cut-off values for the K/iCa ratio were determined to be 4.3 for mortality and 4.1 for active bleeding according to Youden's index. Sensitivity, specificity, and likelihood ratio values for the optimal cut-off value and various K/iCa ratio levels are presented in Table 3.

Patients were divided into 2 groups as those with K/iCa ratio lower and higher than 5.02 according to the K/iCa ratio of 5.02, which has a high specificity value in terms of mortality prediction. Hemodynamic instability was significantly more frequent in the high K/iCa ratio group (88.9% vs. 19.2%, p<0.001). Similarly, active bleeding occurred in 86.1% of patients in the high K/iCa ratio group, compared to 15.1% in the low K/iCa ratio group (p<0.001). RBC suspension was required in 97.2% of the high K/iCa ratio group, whereas 59.6% of the low K/iCa ratio group required transfusions (p<0.001). Additionally, patients with a high K/iCa

Table 1. The endoscopic findings, forest classification and endoscopic interventions

Endoscopic diagnoses	**n (%), all patients**
Gastric erosion/gastritis	497 (71.1)
Gastric/duodenal ulcer	486 (69.5)
Esophageal varices	46 (6.6)
Cancer stomach	44 (6.3)
Dieulafoy's lesion	38 (5.4)
Mallory-Weiss tear	15 (2.1)
Angiodysplasia	1 (0.1)
Active bleeding	131 (18.7)
Forrest classification for gastric/duodenal ulcer, n=486	**n (%)**
1a	10 (2.1)
1b	92 (18.9)
2a	48 (9.9)
2b	142 (29.2)
2c	112 (23)
3	82 (16.9)
Endoscopic intervention	**n (%)**
Sclerotherapy	262 (37.5)
Hemoclip	96 (13.7)
Endoscopic band ligation	42 (6)
Heater probe	7 (1)
Somatostatin	33 (4.7)
Subtotal gastrectomy	7 (1)
Some patients presented with more than 1 endoscopic finding	

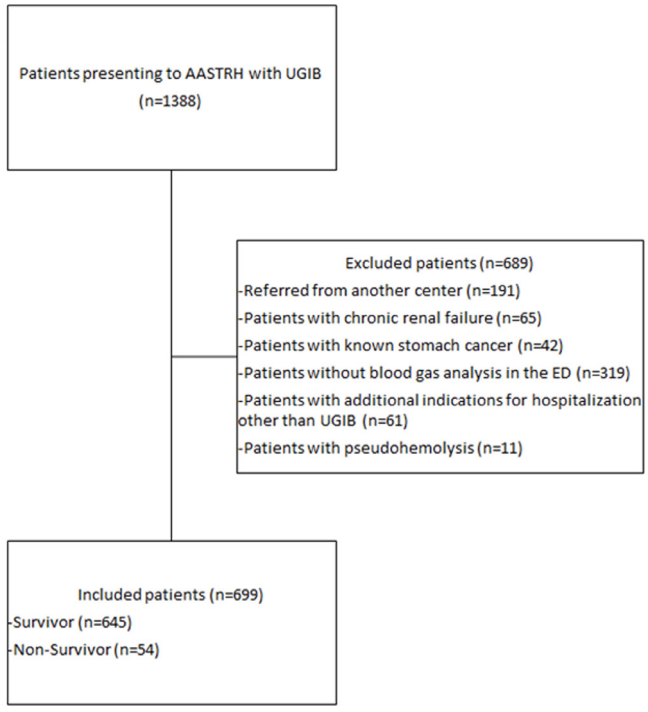


Figure 1. Flowchart showing number of patients of the study

UGIB: Upper gastrointestinal bleeding, AASTRH: American Association for the Surgery of Trauma-Regional Hospital

ratio had a significantly longer hospital stay (median: 6.5 days vs. 4 days, p=0.007). Mortality was significantly higher in the high K/iCa ratio group, occurring in 63.9% of these patients compared to 4.7% in the low K/iCa ratio group (p<0.001). Table 4 presents the comparison of K/iCa ratio with other clinical characteristics.

Discussion

In our study investigating the prognostic value of K value in patients with UGIB, the K/iCa ratio may be a helpful prognostic test for in-hospital mortality and active bleeding in patients with UGIB. Although the likelihood ratios of the best cut-off value of the K/iCa ratio do not have sufficient prognostic value, considering that K and iCa values are inexpensive parameters routinely studied in patients with UGIB, we think that cut-off values with higher specificity or sensitivity, can be helpful instead of the optimal cut-off value of the K/iCa ratio for predicting in-hospital mortality and active bleeding.

In patients presenting to the ED with UGIB, it is difficult to obtain data on bleeding-related factors such as the location, cause, or extent of bleeding. The extent of bleeding should be determined based on the patient's chief complaint, vital signs, and laboratory test results while awaiting the patient's referral for endoscopy.

Table 2. A comparison of patient characteristics with respect to in-hospital mortality (survivor/non-survivor)

	Survivor (n=645)	Non-survivor (n=54)	p value
Age, median, (IQR[1] 25-75)	63 (45-73)	72 (67-81)	**<0.001**
Gender, n (%) Male Female	 443 (68.7) 202 (31.3)	 32 (59.3) 22 (40.7)	 0.154
Comorbidities, n (%) Hypertension Diabetes mellitus CAD[2]/heart failure Liver disease	 330 (51.2) 123 (19.1) 224 (34.7) 26 (4)	 42 (77.8) 20 (37) 31 (57.4) 15 (27.8)	 **<0.001** **0.002** **0.001** **<0.001**
Drugs, n (%) Oral anticoagulants Antiplatelets Acetyl salicylic acid	 84 (13) 46 (7.1) 156 (24.2)	 10 (18.5) 2 (3.7) 13 (24.1)	 0.256 0.572 0.985
Hemodynamic instability, n (%)	107 (16.6)	52 (96.3)	**<0.001**
Median, (IQR 25-75) K[3] (mmol/L) iCa[4] (mmol/L) K/iCa ratio Lactate (mmol/L) Hemoglobin (g/dL) Albumin (g/L) INR[5]	 4.10 (3.78-4.34) 1.07 (1.01-1.12) 3.85 (3.43-4.24) 1.75 (1.21-2.39) 9.7 (8.00-12.4) 36 (33-39) 1.1 (1.04-1.21)	 4.69 (4.18-5.20) 0.96 (0.92-1.00) 4.85 (4.40-5.28) 3.94 (2.82-4.95) 7.25 (5.97-8.22) 27 (24-29) 1.35 (1.20-1.67)	 **<0.001** **<0.001** **<0.001** **<0.001** **<0.001** **<0.001** **<0.001**
Endoscopic diagnoses, n (%) Gastric/duodenal ulcer Esophageal varices Cancer stomach Dieulafoy's lesion	 443 (68.7) 29 (4.5) 34 (5.3) 33 (5.1)	 43 (79.6) 17 (31.5) 10 (18.5) 5 (9.3)	 0.093 **<0.001** **0.001** 0.205
Active bleeding	89 (13.8)	42 (77.8)	<0.001
Need to repeat endoscopy	83 (12.9)	14 (25.9)	0.008
Blood transfusion, n (%) RBC[6] suspension Fresh frozen plasma PCC[7]	 376 (58.3) 58 (9) 19 (2.9)	 54 (100) 39 (72.2) 4 (7.4)	 **<0.001** **<0.001** 0.094
Rockall score stratification Low-risk Moderate-risk High-risk	 537 (83.3) 87 (13.5) 21 (3.3)	 2 (3.7) 37 (68.5) 15 (27.8)	 **<0.001**

IQR: Interquartile range, CAD[2]: Coronary artery disease, K[3]: Potassium, iCa[4]: Ionized calcium, INR[5]: International normalized ratio, RBC[6]: Red blood cell, PCC[7]: Prothrombin complex concentrate

In EDs, especially in developing countries, blood gas analysis is advantageous because it provides results in a much shorter time than other laboratory tests and is inexpensive. K and iCa are parameters that can be obtained by blood gas analysis.

Approximately 45% of calcium, a divalent cation found both inside and outside the cell, is biologically active and ionized, while 55% is bound to proteins such as albumin and citrate. Changes in serum levels of these proteins can lead to imbalances in total body stores and serum calcium levels (4,5,11). Although hypocalcemia in bleeding trauma patients is thought to be related to citrate infusion with blood products, recent studies have shown that most trauma patients have calcium loss prior to blood transfusion due to blood loss, and this is exacerbated by citrate from transfusion (11,12). Furthermore, hypocalcemia on admission has been reported to be a predictor of the need for massive and repeated transfusion (12-14). The same is true for patients with GI bleeding experiencing acute blood loss (1,3).

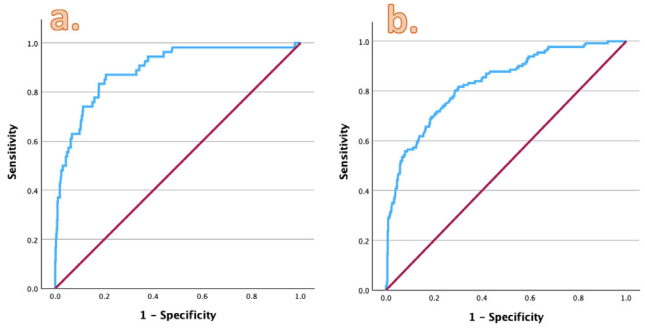


Figure 2. a) ROC analysis to determine K/iCa ratio threshold between non-survivor and survivor groups b) ROC analysis to determine the K/iCa ratio threshold between those with and without active bleeding

K: Potassium, iCa: Ionized calcium

In our study, low iCa was associated with mortality, similar to findings reported in the literature.

Serum K levels in patients with UGIB can vary for several reasons. Serum K levels may decrease due to hemorrhage or increase as a result of the development of prerenal azotemia due to volume depletion in large hemorrhage, direct tissue damage, and K absorption due to digestion of blood from the GI tract (6,7,15). In a retrospective study of trauma patients requiring massive transfusion, the relationship between the K/iCa ratio and mortality was evaluated. The best cut-off value for the K/iCa ratio to distinguish between living and deceased patients was found to be 5.07, and the sensitivity and specificity of this value were reported to be 63.2% and 77.6%, respectively. Patients with a K/iCa ratio above the cut-off had a 4-fold increased risk of death [hazard ratio (HR): 3.97, 95% CI: 1.89-8.32, p<0.001] (7). Another prospective study of trauma patients requiring massive transfusion examined the relationship between the K/iCa ratio measured within 1 hour of hospital admission and mortality. When survivors were compared with deceased patients, a significantly higher mean K/iCa ratio was found (p<0.01). Multivariate logistic regression analysis showed that the total number of blood products was associated with higher K/iCa (odds ratio: 1.02; 95% CI: 1.01-1.04, p=0.01). Cox regression analysis showed a significant association between K/iCa and mortality (HR: 4.12, 95% CI: 1.89-8.96, p<0.001) (16). Based on these studies, we compared the K/iCa ratio with variables such as mortality, transfusion requirement, and hospitalization status in our study of patients with UGIB. In the ROC analysis, we used the cut-off value of the K/iCa ratio between the deceased and living patient groups as 5.02, and the sensitivity and specificity of this value were 42% and 98%, respectively. In addition, hemodynamic instability, active bleeding, and number of blood products

Table 3. The prognostic values for different K/iCa ratio levels to prediction of mortality and active bleeding in patients with UGIB

	AUC[1] (95% CI)	Cut-off value	Sensitivity (95% CI)	Specificity (95% CI)	PLR[2] (95% CI)	NLR[3] (95% CI)	Accuracy (95% CI)
		5.02	42.5 (29.2 to 56.7)	97.9 (96.5 to 98.9)	21.1 (11.3 to 39.1)	0.59 (0.47 to 0.74)	93.7 (91.6 to 95.3)
For mortality	0.892 (0.845-0.939)	4.3*	87 (75.1 to 94.6)	79.3 (76 to 82.4)	4.22 (3.51 to 5.07)	0.16 (0.08 to 0.32)	79.9 (76.8 to 82.8)
		3.88	98.1 (90.1 to 99.9)	52.5 (48.3 to 56.1)	2.06 (1.89 to 2.25)	0.04 (0.01 to 0.28)	55.7 (52 to 59)
		4.73	61.1 (46.8 to 74)	93.3 (91.1 to 95)	9.17 (6.41 to 13.1)	0.42 (0.3 to 0.59)	90 (88.4 to 92.8)
For active bleeding	0.832 (0.792-0.871)	4.1*	80.1 (72.2 to 86.6)	71.3 (67.3 to 74.9)	2.79 (2.39 to 3.26)	0.28 (0.2 to 0.4)	72 (69.5 to 76.2)
		3.64	98 (90 to 99)	36.9 (33 to 40.7)	1.56 (1.46 to 1.67)	0.05 (0.01 to 0.35)	41.6 (37.9 to 45.3)

*The best cut-off value was calculated according to the Youden index. AUC[1]: Area under curve, PLR[2]: Positive likelihood ratio, NLR[3]: Negative likelihood ratio, UGIB: Upper gastrointestinal bleeding, CI: Confidence interval

Table 4. Comparison of patient characteristics with respect to the K/iCa ratio			
	K/iCa[1] ratio<5.02 (n=663)	K/iCa ratio>5.02 (n=36)	p value
Hemodynamic instability, n (%)	127 (19.2)	32 (88.9)	**<0.001**
Active bleeding, n (%)	100 (15.1)	31 (86.1)	**<0.001**
Need to repeat endoscopy, n (%)	88 (13.3)	9 (25)	**0.077**
Blood transfusion, n (%) RBC[2] suspension	 395 (59.6)	 35 (97.2)	 **<0.001**
Somatostatin, n (%)	24 (3.6)	9 (25)	**<0.001**
Rockall score stratification, n (%) Low-risk Moderate-risk High-risk	 532 (80.2) 105 (15.8) 26 (3.9)	 7 (19.4) 19 (52.8) 10 (27.8)	 **<0.001**
Median, (IQR[3] 25-75) Number of RBC suspensions Length of hospital stay Length of ICU[4] stay	 2 (0-3) 4 (2-6) 3 (3-6)	 6 (4-8) 6.5 (2-16.25) 5 (2-17.5)	 **<0.001** **0.007** 0.140
Mortality, n (%) Non-survivor	 31 (4.7)	 23 (63.9)	 **<0.001**
K/iCa[1] ratio: Potassium/ionized calcium ratio, RBC[2]: Red blood cell, IQR[3]: Inter quartile range, ICU[4]: Intensive care unit			

administered and hospital days were found to be higher in the higher K/iCa ratio group ($p<0.05$). These studies have shown that K levels increase and iCa levels decrease during acute blood loss. In trauma patients, K levels increase proportionally to tissue damage. In UGIB patients, there is a linear relationship between the amount of bleeding and K levels. This is because bleeding causes tissue damage, blood is hemolyzed in the GII tract, and K is reabsorbed. Particularly in elderly patients, acute deterioration of renal function may occur when significant volume depletion is added to an already low glomerular filtration rate. This may also contribute to elevated K levels. In the setting of acute blood loss, we found a decrease in iCa levels in our study and in previous studies. It is known that K levels increase and calcium levels decrease, due to citrate binding to calcium, especially when blood products are transfused to hemodynamically unstable patients with active bleeding. For these reasons, we suggest that a high K/iCa ratio may be a helpful marker for the determination of mortality risk in UGIB patients.

Study Limitations

This study was conducted retrospectively at a single health center. Some symptoms or medical histories may not have been accurately recorded due to missing information. The sample size of our study is limited, and we may not have sufficient statistical power for some subgroup analyses. The cut-off value for the K/iCa ratio used in our study may be different in other studies. Patients with chronic renal failure who were already expected to have high K levels and patients with other indications for hospitalization were not included in the study. If these patients had been included, the mortality rate in patients with a high K/iCa ratio may have changed. In addition, only in-hospital mortality was analyzed. The study did not examine factors that influence long-term mortality.

Conclusion

In patients presenting to the ED with suspected UGIB, assessment of the K/iCa ratio using blood gas analysis may help to predict the risk of active bleeding and in-hospital mortality as a rapid, inexpensive and practical assessment method. Particularly, a cut-off value of 5.02 demonstrated high specificity in predicting mortality and should be considered in clinical practice. Furthermore, a high K/iCa ratio has been observed to be associated with adverse clinical outcomes such as hemodynamic instability, active bleeding, need for erythrocyte suspension, and prolonged hospital stay. Therefore, patients with a high K/iCa ratio may be candidates for early endoscopy and closer hemodynamic monitoring. However, these findings need to be confirmed by larger prospective studies before clinical application.

Ethics

Ethics Committee Approval: The study was approved by the local ethics committee Atatürk Sanatorium Training and Research Hospital (decision number: 20-KAEK15/2854, date: 27.12.2023).

Informed Consent: Retrospective study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: M.U., H.Ö.O., E.E., Concept: H.Ö.O., Y.Ç., E.E., Design: S.K.Ç., Y.Ç., E.E., Data Collection or Processing: H.Ö.O., S.A., Analysis or Interpretation: E.E., H.Ö.O., S.K.Ç., Literature Search: S.A., H.Ö.O., Writing: H.Ö.O., 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. Korytny A, Mazzawi F, Marcusohn E, Klein A, Epstein D. Admission hypocalcemia and the need for endoscopic and clinical interventions among patients with upper gastrointestinal bleeding. Eur Surg Res. 2023;64:398-405.
2. Emektar E, Dağar S, Çorbacıoğlu ŞK, Uzunosmanoğlu H, Uzman M, Karaatlı RH, et al. Predictors of mortality in geriatric patients with upper gastrointestinal bleeding. Eurasian J Emerg Med. 2020;19:197-202.
3. Korytny A, Klein A, Marcusohn E, Freund Y, Neuberger A, Raz A, et al. Hypocalcemia is associated with adverse clinical course in patients with upper gastrointestinal bleeding. Intern Emerg Med. 2021;16:1813-22.
4. Robinson A, Rech MA, DeChristopher PJ, Vaughn A, Rubino J, Bannister E, et al. Defining the optimal calcium repletion dosing in patients requiring activation of massive transfusion protocol. Am J Emerg Med. 2023;70:96-100.
5. Wynne Z, Falat C. Disorders of calcium and magnesium. Emerg Med Clin North Am. 2023;41:833-48.
6. Chen L, Zheng H, Wang S. Prediction model of emergency mortality risk in patients with acute upper gastrointestinal bleeding: a retrospective study. PeerJ. 2021;9:e11656.
7. Ninokawa S, Tatum D, Toraih E, Nordham K, Ghio M, Taghavi S, et al. Elevated K/iCa ratio is an ancillary predictor for mortality in patients with severe hemorrhage: a decision tree analysis. Am J Surg. 2022;223:1187-93.
8. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open. 2016;6:e012799.
9. Gilbert EH, Lowenstein SR, Koziol-McLain J, Barta DC, Steiner J. Chart reviews in emergency medicine research: where are the methods? Ann Emerg Med. 1996;27:305-8.
10. Çorbacıoğlu ŞK, Aksel G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: a guide to interpreting the area under the curve value. Turk J Emerg Med. 2023;23:195-8.
11. Wray JP, Bridwell RE, Schauer SG, Shackelford SA, Bebarta VS, Wright FL, et al. The diamond of death: hypocalcemia in trauma and resuscitation. Am J Emerg Med. 2021;41:104-9.
12. Magnotti LJ, Bradburn EH, Webb DL, Berry SD, Fischer PE, Zarzaur BL, et al. Admission ionized calcium levels predict the need for multiple transfusions: a prospective study of 591 critically ill trauma patients. J Trauma. 2011;70:395-7.
13. Kronstedt S, Roberts N, Ditzel R, Elder J, Steen A, Thompson K, et al. Hypocalcemia as a predictor of mortality and transfusion. A scoping review of hypocalcemia in trauma and hemostatic resuscitation. Transfusion. 2022;62 (Suppl 1):158-66.
14. Wang Z, Chen X, Chen Y, Yang L, Wang H, Jiang W, et al. Low serum calcium is associated with perioperative blood loss and transfusion rate in elderly patients with hip fracture: a retrospective study. BMC Musculoskelet Disord. 2021;22:1025.
15. Emektar E. Acute hyperkalemia in adults. Turk J Emerg Med. 2023;23:75-81.
16. Gagen B, Duchesne J, Ghio M, Duplechain A, Krakosky D, Simpson JT, et al. Prospective examination of the K/iCa ratio as a predictor for mortality in severe hemorrhage. Shock. 2024;61:34-40.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.93276
Eurasian J Emerg Med. 2025;24(3): 183-9

Healthcare Under Attack

Hasan Demirbaş[1], Ekim Sağlam Gürmen[2], Mustafa Yorgancıoğlu[3]

[1]Gölcük Necati Çelik State Hospital, Clinic of Emergency Department, Kocaeli, Türkiye
[2]Manisa Celal Bayar University Faculty of Medicine, Department of Emergency Department, Manisa, Türkiye
[3]İzmir Torbalı State Hospital, Clinic of Emergency Department, İzmir, Türkiye

Abstract

Aim: Social media triggers an increase in violent incidents in all occupational groups, especially in the health sector. The aim of our study is to identify preventable causes of violence against healthcare workers and reveal the measures that can be taken.

Materials and Methods: Survey questions that would reveal the biopsychosocial effects of violent events were asked to healthcare professionals, and the answers were recorded.

Results: A total of 1006 healthcare workers were included in the study. While 87.8% of the healthcare workers stated that they were exposed to verbal violence, 17.8% indicated that they were exposed to physical violence. While 79.9% of the female healthcare workers stated that violent news on social media increases violence in healthcare, the majority of the healthcare workers stated that the news is not reflected objectively. Nurses experienced the highest rate of verbal violence at 90.2% and applied code white (code violet) the most frequently at a rate of 36.6%.

Conclusion: It is noticeable that the number of cases of violence in the health sector has increased due to unconfirmed news and misinformation on social media. Almost all healthcare workers stated that they were subjected to violence at least once in their working lives, and that they wanted to change their field of work. To improve working conditions and prevent violence, the competent authorities should immediately implement the necessary penal actions that are also deterrent.

Keywords: Healthcare, social media, violence, violent news

Introduction

Violence has existed since the beginning of human history and affects individuals, societies, and all social professions in various ways and degrees according to cultural characteristics. According to the World Health Organization, violence it is defined as "the threat or deliberate use of force against oneself, another person or a group resulting in injury, death, psychological harm, developmental delay or deprivation" (1). Violence in healthcare is defined as "coming from the patient, patient's relatives or any other individual, posing a risk to the healthcare professional: situations consisting of threatening behavior, verbal threats, economic abuse, physical assault, and sexual assault" (2).

Exposure to violence has been increasing in recent years for all professional groups, especially healthcare workers and physicians, and constitutes a serious health problem. Healthcare workers are at risk for violence all over the world. Between 8% and 38% of employees are exposed to physical violence. In addition, verbal violence and threat rates are much higher than previously reported populations (3). It has been determined that people working in the health sector have a 16-fold increased risk of violence compared to those working in other sectors (4). In addition, one study showed that 25% of all incidents of violence occur in the health sector. It has been stated that 50% of healthcare workers are exposed to violence in different dimensions, from physical violence to psychosocial violence (5). Some factors that increase violence in health institutions are 24-hour uninterrupted service, family members' inability to cope with stress, extended waiting time due to high patient volume, patients not benefiting from care services as much as they want, and insufficient healthcare and security personnel (6,7).

Corresponding Author: Ekim Sağlam Gürmen MD, Manisa Celal Bayar University Faculty of Medicine, Department of Emergency Department, Manisa, Türkiye
E-mail: ekimdr@hotmail.com **ORCID ID:** orcid.org/0000-0002-8672-6181
Cite this article as: Demirbaş H, Sağlam Gürmen E, Yorgancıoğlu M. Healthcare under attack. Eurasian J Emerg Med. 2025;24(3): 183-9.

Received: 11.02.2025
Accepted: 23.03.2025
Epub: 25.04.2025
Published: 10.09.2025

Regarding occupational groups, general practitioners are the most frequently exposed to violence, followed by assistant doctors and nurses. When the experience of the employees was examined, the risk of exposure to violence was found to be high, especially in the first 10 years of working life (8). When people who have been subjected to violence are examined, women are found to have a higher relative risk of being exposed to violence (9).

The frequency of violence in the workplace is not clearly known due to the deficiencies and disruptions in the notifications, which are limited only to self-reporters. One study in our country determined that 67% of healthcare workers, 62% of physicians, and 60% of emergency service workers did not report violence or register any complaints (10). When examining violence toward healthcare workers, the majority of people who exhibited violence were between the ages of 21 and 30 years and the relatives of the patients were more inclined to violence than the patients (11).

Occasionally, biased news targeting healthcare professionals is produced in the media (12). The blatant encouragement of violence and the presentation of violence as a search for rights also increase violence against healthcare workers. Presenting a complication occurring during healthcare services as a medical error without adequate investigation also increases the incidence of violence (13).

In our country, as in the rest of the world, some regulations have been made to prevent violence in healthcare. Procedures and principles regarding legal aid to healthcare workers subjected to violence have been established. In this context, the Code White Call system has been implemented, which aims to enable communication in case of a risk of violence, save time for intervention, and ensure employee safety (14).

The aim of our study is to identify preventable causes of violence in healthcare workers and reveal the measures that can be taken. In addition, we investigated the level of exposure to violence and the importance of violent news via social media on the biopsychosocial lives of healthcare professionals working in a tertiary university hospital.

Materials and Methods

This study was conducted under the leadership of the research team between 01.11.2023 and 01.05.2024 at Manisa Celal Bayar University Faculty of Medicine Hospital, Emergency Medicine Clinic Seminar Hall. Healthcare workers in this study were asked survey questions created by the researchers. The survey lasted approximately 5 minutes. Answers to the questionnaire form were recorded. Informed consent was obtained from healthcare workers. The confidentiality of each participant who filled out the questionnaire was ensured. This survey was conducted according to the Personal Data Protection Law. The survey questions, created by the researchers and designed to take approximately 5 minutes, were administered to volunteer participants under the guidance of the research team at Manisa Celal Bayar University. The responses given to the survey form were recorded in the study form.

Ethics approval was obtained from Manisa Celal Bayar University Non-Interventional Clinical Research Ethics Committee with (decision number: 20.478.486-2071, date: 08.11.2023).

Criteria for Inclusion in the Study:

- Healthcare workers including doctors, nurses, midwives, and health technicians at Manisa Celal Bayar University Faculty of Medicine Hospital

Criteria for Exclusion from the Study:

- Healthcare workers who did not give informed voluntary consent
- Incomplete questionnaires

Statistical Analysis

All volunteer healthcare workers at Manisa Celal Bayar University Faculty of Medicine Hospital were included in the study. All statistical analyses were performed using SPSS for Windows, version 11.5 (SPSS Inc.). The Shapiro-Wilk test was used to assess the normality assumption for the variables “age” and “working year”. In the descriptive statistics, median (minimum-maximum) values were used for “age” and “working year” because they were non-normally distributed. Numbers (n) and percentages (%) were used for categorical variables.

The comparisons between survey questions and demographic variables (gender and occupational groups) were examined using the Pearson chi-square or Fisher’s exact test A p value less than 0.05 was considered statistically significant.

Results

A total of 1006 people participated in this study. Of those, 760 (75.5%) were female. The median age of the healthcare workers was 28 (17-68) years. Of the healthcare workers, 413 (41.1%) were doctors, 336 (33.4%) were nurses, 218 (21.7%) were health technicians, and 39 (3.9%) were midwives. The median duration of employment was 5 years (ranging from 1 to 42 years).

Regarding exposure to violence, 883 (87.8%) healthcare workers stated that they were exposed to verbal violence, while 179 (17.8%) indicated exposure to physical violence. The rate of those

who triggered code white (code violet) in cases of violence was 30.5% (n=307).

When we examined the effects of violence on social media of the healthcare workers, 42.1% stated that their work motivation was affected, and 35.2% stated that their career plans were negatively affected.

When we asked the healthcare workers if they had ever considered living or working abroad, 85.2% stated that they had.

While 56.7% stated they are members of a professional organization, association, and/or health union, 79.5% of the healthcare workers stated that professional organizations, associations, and health unions do not have sufficient initiatives and efforts to prevent violence.

The rates of healthcare workers stating that X-ray devices and security guards are insufficient to prevent violence were 76.3% and 90.7%, respectively, (Table 1).

When we compared the exposure of healthcare workers to violence, exposure to verbal violence was more frequent among females than males (88.9% vs. 84.1%, p=0.046).

Table 1. Healthcare workers views on violence in healthcare (n=1006)

Survey questions	n (%)
Have you ever been exposed to verbal violence at some point in your working life?	
Yes	883 (87.8)
No	123 (12.2)
Have you ever been exposed to physical violence at some point in your working life?	
Yes	179 (17.8)
No	827 (82.2)
If you have been exposed to violence before, have you ever applied code white for violence?	
Yes	307 (30.5)
No	699 (69.5)
Do the incidents of violence in health on social media negatively affect your work motivation?	
It does not affect at all	5 (0.5)
Infrequently affects	36 (3.6)
Sometimes it affects	176 (17.5)
It often affects	424 (42.1)
Constantly influencing	365 (36.3)
Do you think that the violent news in health on social media plays a role in the increase in violence in health?	
Has an effect	777 (77.2)
No effect	89 (8.8)
I have no idea	140 (14.0)
Do you think that the violent news in health on social media is published objectively and accurately?	
Yes	84 (8.4)
No	763 (75.8)
I have no idea	159 (15.8)
Are you a member of a professional organization, association, and/or health union?	
Yes	570 (56.7)
No	436 (43.3)
Do you think professional organizations, associations, and health unions have enough initiative and effort to prevent violence?	
Yes	52 (5.2)
No	800 (79.5)
I have no idea	154 (15.3)
Do you think that security personnel in hospitals are effective in preventing violence in healthcare?	
Yes	55 (5.5)
No	912 (90.7)
I have no idea	39 (3.8)
Do you think that X-ray security devices placed at hospital entrances are effective in preventing violence in healthcare?	
Yes	188 (18.7)
No	768 (76.3)
I have no idea	50 (5.0)

When we examined the influence of gender-based violence, the impact on work motivation was more prominent in females than in males (44.1% vs. 36.2%, p=0.007). When we asked about changing their professions, the desire to change professions was more frequent in females than males (22.5% vs. 16.3%, p=0.039).

Answers to the question "Do you think that the violent news on social media has a role in the increase in violence incidents in healthcare?" indicated that more females than males believe this to be the case, with 79.9% vs. 69.1% agreeing, respectively (p=0.002) (Table 2).

There was no statistically significant difference between occupational groups in terms of code white's ability to effectively address violent incidents, although negative opinions were dominant. A statistically significant difference was found in terms of application of "code white" (p=0.024). Of the nurses, 36.6% declared that they applied code white.

When we examined the effect of work motivation affected by violence among occupational groups, a statistically significant difference was observed between the groups (p=0.001). 28.1%

Table 2. Comparison of healthcare workers by gender (n (%)			
Survey questions	**Male (n=246)**	**Female (n=760)**	**p value**
Have you ever been exposed to verbal violence at some point in your working life?			
Yes	207 (84.1)[a]	676 (88.9)[b]	0.046*
No	39 (15.9)[a]	84 (11.1)[b]	
Have you ever been exposed to physical violence at some point in your working life?			
Yes	45 (18.3)	134 (17.6)	0.814*
No	201 (81.7)	626 (82.4)	
Do the incidents of violence in health on social media negatively affect your work motivation?			
It does not affect at all	3 (1.1)[a]	2 (0.2)[a]	0.007**
Infrequently affects	12 (4.9)[a]	24 (3.2)[a]	
Sometimes it affects	57 (23.2)[a]	119 (15.7)[b]	
It often affects	89 (36.2)[a]	335 (44.1)[b]	
Constantly influencing	85 (34.6)[a]	280 (36.8)[a]	
Do you think that the news of violence in health on social media plays a role in the increase in violence in health?			
Has an effect	170 (69.1)[a]	607 (79.9)[b]	0.002*
No effect	28 (11.4)[a]	61 (8.0)[a]	
I have no idea	48 (19.5)[a]	92 (12.1)[b]	
Have you ever considered settling or working abroad after the news of violence in health on social media?			
I have never considered	22 (8.9)[a]	127(16.7)[b]	0.006*
I have considered infrequently	29 (11.8)[a]	89 (11.7)[a]	
Sometimes, I consider	54 (22.0)[a]	196 (25.8)[a]	
I often consider	62 (25.2)[a]	168 (22.1)[a]	
I have always considered	79 (32.1)[a]	180 (23.7)[b]	

*Pearson chi-square **Fisher's exact test. a,b: Different letter indices indicate statistically significant differences between groups according to pairwise comparisons

of the doctors stated that they were affected. A statistically significant difference was found between occupational groups regarding the change of profession due to the violence incidents (p=0.001). Of the doctors, 32.9% answered this question as "Sometimes, I think".

A statistically significant difference was observed between professional groups when they were asked whether they thought of settling or working abroad after the news of violence on social media (p=0.001). Of the doctors, 30.0% stated that they constantly think about this situation.

Regarding the effectiveness of X-ray security devices placed at hospital entrances in preventing violence in healthcare, a statistically significant difference was detected between professional groups (p=0.028). Additionally, 33.3% of midwives stated that X-ray security devices are effective in this regard.

When opinions on the effects of professional organizations, associations, or unions (p=0.001) and security personnel (p=0.002) in preventing violence were evaluated, a statistically significant difference was found between society members and non-members. Eighty-five point six percent of the society members and 71.6% of the non-members stated that these organizations did not have sufficient initiatives to prevent violence. In addition, 93% of those with membership and 87.6% without membership stated that security personnel are ineffective in preventing violence.

Discussion

In this study, 87.8% of the healthcare workers stated that they were exposed to verbal violence, and 17.8% were exposed to physical violence. In a study conducted across public hospitals in Palestine, 59.6% of healthcare workers stated that they were exposed to verbal violence, and 20.8% indicated that they were exposed to physical violence. In contrast, in a study conducted on emergency service workers in Palestine, 71.2% of the healthcare workers stated that they were exposed to verbal violence, and 35.6% indicated that they were exposed to physical violence (15,16). In a study conducted in China, it was observed that 92.7% of physicians were exposed to verbal violence (17). In a multicenter study conducted in Türkiye, it was stated that 92.6% of women and 87.5% of men were exposed to verbal violence (18). Our study is consistent with the literature, indicating that the most prevalent form of violence is verbal violence.

In our study, when we looked at the exposure of occupational groups to violence, there was no statistically significant difference in exposure to verbal violence. Nurses experienced the highest exposure to verbal violence at 90.2%, followed by doctors, midwives, and health technicians, respectively. Nurses (23.5%) and health technicians (24.8%) were exposed to physical violence more frequently than doctors and midwives. In a study conducted in Edirne, nurses were the profession that were exposed to the most violence throughout their professional life, with 85% (19). In this respect, while our study is consistent with the literature, some studies state that the physician group is exposed to more violence than the nurse group in England and Palestine (15,20). Reasons such as the field differences of the employees participating in the

Table 3. Comparison of healthcare workers by occupational groups n (%)					
Survey questions	**Doctors (n=413)**	**Nurses (n=336)**	**Midwives (n=39)**	**Health technicians (n=218)**	**p value**
Have you ever been exposed to verbal violence at some point in your working life?					0.132*
Yes	364 (88.1)	303(90.2)	34 (87.2)	182 (83.5)	
No	49 (11.9)	33 (9.8)	5 (12.8)	36 (16.5)	
Have you ever been exposed to physical violence at some point in your working life?					0.001*
Yes	43 (10.4)[a]	79 (23.5)[b]	3 (7.7)[a,b]	54 (24.8)[b]	
No	370 (89.6)[a]	257 (76.5)[b]	36 (92.3)[a,b]	164 (75.2)[b]	
If you have been exposed to violence before, have you ever applied code white for violence?					0.024*
Yes 112 (27.1)[a]	123 (36.6) b	9 (23.1) a,b	63 (28.9)[a,b]		
No	301 (72.9)[a]	213 (63.4) b	30 (76.9)[a,b]	155 (71.1)[a,b]	
Have you ever considered settling or working abroad after the violent news in health on social media?					0.001*
I have never considered	37 (9.0)[a]	58 (17.3)[b]	9 (23.1)[b]	45 (20.6)[b]	
I have considered infrequently	46 (11.1)[a]	40 (11.9)[a]	8 (20.5)[a]	24 (11.0)[a]	
Sometimes, I consider	102 (24.7)[a]	92 (27.4)[a]	6 (15.4)[a]	50 (22.9)[a]	
I often consider	104 (25.2)[a]	74 (22.0)[a]	11 (28.2)[a]	41 (18.8)[a]	
I have always considered	124 (30.0)[a]	72 (21.4)[b]	5 (12.8)[a,b]	58 (26.6)[a,b]	
Do you think that the health violence news on social media is published objectively and accurately?					0.001*
Yes	21 (5.1)[a]	39 (11.6)[b]	5 (12.8)[a,b]	19 (8.7)[a,b]	
No	337 (81.6)[a]	248 (73.8)[a,b]	22 (56.4)[b]	156 (71.6)[b]	
I have no idea	55 (13.3)[a]	49 (14.6)[a,b]	12 (30.8)[b]	43 (19.7)[a,b]	
Are you a member of any professional organization, association and/or health union?					0.001*
Yes	160 (38.7)[a]	242 (72.0)[b,c]	33 (84.6)[b]	135(61.9)[c]	
No	253 (61.3)[a]	94 (28.0)[b,c]	6 (15.4)[b]	83 (38.1)[c]	
Do you think professional organizations, associations, and health unions have enough initiative and effort to prevent violence?					0.003*
Yes	29 (7.0)[a]	14 (4.2)[a]	2 (5.1)[a]	7 (3.2)[a]	
No	304 (73.6)[a]	287 (85.4)[b]	34 (87.2)[a,b]	175 (80.3)[a,b]	
I have no idea	80 (19.4)[a]	35 (10.4)[b]	3 (7.7)[a,b]	36 (16.5)[a,b]	
*Pearson chi-square. a,b,c: Different letter indices indicate statistically significant differences between groups according to the pairwise comparisons					

survey, the role of the profession in the country, and differences in experience, may have led to different results among the studies.

Many studies have determined that female health workers are exposed to violence more than male health workers, both in Türkiye and abroad (21). In our study, in line with the literature, verbal violence was significantly more frequent in female healthcare workers than in male healthcare workers (88.9% vs. 84.1%, p=0.046).

When examining complaints about violent acts, and the implementation of code white in our study, there was no difference between the genders. However, it was determined that the nurses were the group that applied the code white with the highest rate, at 36.6% (p=0.024). It is seen that reports on the incidence of violence are limited, regardless of sex and occupational group. When the studies in the literature are examined, it appears that a small number of violent incidents in healthcare are reported (22-24). The code white data, the

people who reported the violence were mainly physicians, and our study, therefore, differs from the literature (25,26).

40.9% of the healthcare workers stated that they were often negatively affected psychologically due to news about violence in healthcare on social media; 62.3% indicated that their professions were constantly devalued; and 78.4% stated that this news negatively affected their work motivation. The literature indicates that there is a link between exposure to violence in the workplace and emotional exhaustion. There is a decrease in job satisfaction and motivation among employees who are exposed to violence and symptoms of posttraumatic stress disorder may occur (27-29).

In previous survey studies conducted in the USA, patients, their relatives, and healthcare professionals stated that they felt safe due to the X-ray devices placed at the hospital entrances (30). Again, among the measures that can be taken to solve violent incidents are 24-hour security, security doors, security cameras, and panic alarms (31). In a study conducted in İzmir, 91.6% of health workers stated that security guards were inadequate in preventing violence (32). In our study, 76.3% of the healthcare workers thought that X-ray devices and 90.7% thought that security guards were ineffective and insufficient in preventing violence. Regardless of the occupational group, most of the employees stated that security guards were ineffective in preventing violence. The ineffectiveness of X-ray devices, contrary to what is stated in the literature, can be attributed to several factors. These include the absence of devices at every entrance, underutilization of available devices, and limited availability of devices in hospitals.

In our study, 56.7% of healthcare workers are members of a professional organization; however, only 21.5% of them believe that these organizations make sufficient efforts to prevent violence.

Violence in health is a problem that our country and the whole world face, and it is growing daily. The measures taken to ensure that healthcare workers work in a peaceful and safe environment are insufficient. In many hospitals, security personnel are insufficient, and the interventions have been delayed due to the limitations on authority.

Study Limitations

Since this study was conducted in a single center, the data do not cover the entire region. Our hospital is a tertiary university hospital, so the severity of the patients may affect the incidences of violence. For this reason, more comprehensive and multicenter studies are needed.

Conclusion

People's use of social media is increasing every day. For this reason, access to information is very rapid. Often, the news is not checked to see whether it is true or not. With fake news and misinformation, people may turn against healthcare workers through social media. Therefore, this increases the incidence of violence.

Considering all of this, it is necessary to implement legal deterrents to reduce violence in healthcare. In this way, healthcare workers who experience incident should be encouraged to report via code white. Professional organizations should work more on this issue, and support their members. Black propaganda against health workers and news of violence on social media should be identified, and necessary punitive and deterrent actions should be taken.

Ethics

Ethics Committee Approval: Ethics approval was obtained from Manisa Celal Bayar University Non-Interventional Clinical Research Ethics Committee with (decision number: 20.478.486-2071, date: 08.11.2023).

Informed Consent: Informed consent was obtained from healthcare workers.

Footnotes

Author Contributions

Surgical and Medical Practices: H.D, E.S.G., Concept: H.D, E.S.G., M.Y., Design: H.D, E.S.G., M.Y., Data Collection or Processing: H.D, E.S.G., M.Y., Analysis or Interpretation: H.D, E.S.G., M.Y., Literature Search: H.D, E.S.G., M.Y., Writing: H.D, E.S.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. World report on violence and health: summary. Geneva, World Health Organization. 2002. Available from: https://www.who.int/publications/i/item/9241545615
2. Saines JC. Violence and aggression in A&E: recommendations for action. Accid Emerg Nurs. 1999;7:8-12.
3. Violence against health workers, World Health Organisation (WHO). Available from: https://www.who.int/activities/preventing-violence-against-health-workers
4. Kingma M. Workplace violence in the health sector: a problem of epidemic proportion. Int Nurs Rev. 2001;48:129-30.
5. Celbiş B, Özaydın Ö. Examının the relationshıp between violence against healthcare workers and their levels of exhaustion, confıdence and safety. Ege Academic Review. 2024;24:421-32.

6. Aktuğ K, Hancı H. Acil serviste şiddet tehdidi: hekimin yasal sorumlulukları ve hakları (tıp ve sağlık hukuku) İzmir: Ege Üniversitesi Tıp Fakültesi, Toprak Ofset. Available from: https://www.mevzuatdergisi.com/1999/05a/01.htm
7. Cooper CL, Swanson N. Workplace violence in the health sector state of the art. 2008. Available from: https://www.who.int/docs/default-source/documents/violence-against-health-workers/wvstateart.pdf?sfvrsn=36aae706_2
8. Ayranci U, Yenilmez C, Balci Y, Kaptanoglu C. Identification of violence in Turkish health care settings. J Interpers Violence. 2006;21:276-96.
9. Arnetz J, E Arnetz, B B, Petterson IL. Violence in the nursing profession: occupational and lifestyle risk factors in Swedish nurses. Work & Stress. 1996;10:119-27.
10. Aydın M. Isparta-Burdur sağlık çalışanlarına yönelik şiddet ve şiddet algısı. Türk Tabipleri Birliği, Isparta- Burdur Tabip Odası Başkanlığı. 2008.
11. Erkol H, Gökdoğan MR, Erkol Z, Boz B. Aggression and violence towards health care providers--a problem in Turkey? J Forensic Leg Med. 2007;14:423-8.
12. Ankara Tabip Odası. Hekimlere ve sağlık çalışanlarına yönelik şiddet raporu, Ankara Tabip Odası Yayınları. 2011.
13. Sarcan E. Toplumun sağlık çalışanlarına uygulanan şiddete bakış açısı. Gaziantep Üniversitesi Tıp Fakültesi Acil Tıp Anabilim Dalı Uzmanlık Tezi. 2013.
14. Aktaş E, Aydemir İ. Şiddete maruz kalan sağlık çalışanlarının beyaz kod uygulamasına ilişkin görüşlerinin belirlenmesi. Turk Klin J Health Sci. 2018;3:32-47.
15. Kitaneh M, Hamdan M. Workplace violence against physicians and nurses in Palestinian public hospitals: a cross-sectional study. BMC Health Serv Res. 2012;12:469.
16. Hamdan M, Abu Hamra A. Workplace violence towards workers in the emergency departments of Palestinian hospitals: a cross-sectional study. Hum Resour Health. 2015;13:28.
17. Shi J, Wang S, Zhou P, Shi L, Zhang Y, Bai F, Xue D, Zhang X. The frequency of patient-initiated violence and its psychological impact on physicians in china: a cross-sectional study. PLoS One. 2015;10:e0128394.
18. Aydin B, Kartal M, Midik O, Buyukakkus A. Violence against general practitioners in Turkey. J Interpers Violence. 2009;24:1980-95.
19. Çuvadar A, Ekuklu G. The frequency and causes of violence against doctors and nurses in district hospitals in Edirne City. Anadolu Hemşirelik ve Sağlık Bilimleri Dergisi. 2020;23:212-20.
20. Quine L. Workplace bullying in junior doctors: questionnaire survey. BMJ. 2002;324:878-9.
21. İlhan MN, Özkan S, Kurtcebe ZÖ, Aksakal FN. Gazi Üniversitesi Tıp Fakültesi Hastanesinde çalışan araştırma görevlileri ve intörn doktorlarda şiddete maruziyet ve şiddetle ilişkili etmenler. 2009;28: 15-23.
22. Bahar A, Şahin S, Akkaya Z, Alkayiş M. Acil serviste çalışan hemşirelerin şiddete maruz kalma durumu ve iş doyumuna etkili olan faktörlerin incelenmesi. Psikiyatri Hemşireliği Dergisi. 2015;6:57-64.
23. Demiroğlu T, Kılınç E, Atay E. Sağlık çalışanlarına uygulanan şiddet: Kilis ili örneği. Sağlık Bilimleri Dergisi. 2015;24:49-55.
24. Kahya B, Çetin C, Cebeci F. Acil servis hemşirelerine yönelik şiddet: sistematik derleme. Fırat Tıp Dergisi. 2016;21:170-76.
25. Polat Ö, Çırak M. Sağlıkta şiddetin beyaz kod verileri ile değerlendirilmesi. Bakırköy Tıp Dergisi. 2019;15:388-93.
26. Yılmaz K, Kaya K, Özesen TA, Hilal A. Adana ilinde sağlık çalışanlarının şiddete uğrama sıklığı ve sağlıkta şiddet konusundaki düşünceleri. 2023;28:151-59.
27. Nart, S. İş ortamında şiddet, tükenmişlik ve iş tatmini ilişkileri: sağlık çalışanları üzerinde bir araştırma. Yönetim ve Ekonomi Araştırmaları Dergisi. 2014;12:248-68.
28. Vartia-Väänänen M. Workplace bullying- A study on the work environment, well-being and health, Unpublished PhD Thesis, University of Helsinki Department of Psychology- Finnish Institute of Occupational Health, Helsinki. 2003.
29. Coles J, Koritsas S, Boyle M, Stanley J. GPs, violence and work performance – 'just part of the job?'. Aust Fam Physician. 2007;36:189-91.
30. Meyer T, Wrenn K, Wright SW, Glaser J, Slovis CM. Attitudes toward the use of a metal detector in an urban emergency department. Ann Emerg Med. 1997;29:621-4.
31. Kuhn W. Violence in the emergency department. Managing aggressive patients in a high-stress environment. Postgrad Med. 1999;105:143-8,154.
32. Durak TÇ, Yolcu S, Akay S, Demir Y, Kılıçaslan R, Değerli V, et al. Bozyaka Eğitim ve Araştırma Hastanesi sağlık çalışanlarına hasta ve hasta yakınları tarafından uygulanan şiddetin değerlendirilmesi. Genel Tıp Derg. 2014;24:130-7.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.37108
Eurasian J Emerg Med. 2025;24(3): 190-5

The Role of Serum Beta-Trace Protein in the Diagnosis and Prognosis of Sepsis Patients in the Emergency Department

Mehmet Yortanlı[1], Ramazan Köylü[2], Öznur Köylü[3], Betül Çiğdem Yortanlı[4], Başar Cander[5]

[1]Konya Numune Hospital, Clinic of Emergency Medicine, Konya, Türkiye
[2]Konya City Hospital, Clinic of Emergency Medicine, Konya, Türkiye
[3]Konya City Hospital, Clinic of Biochemistry, Konya, Türkiye
[4]Konya City Hospital, Clinic of Internal Medicine, Konya, Türkiye
[5]Bezmialem Vakıf University Faculty of Medicine, Department of Emergency Medicine, İstanbul, Türkiye

Abstract

Aim: This study aimed to evaluate the diagnostic and prognostic significance of serum beta-trace protein (BTP) levels in patients diagnosed with sepsis in the emergency department.

Materials and Methods: This prospective, single-center, observational clinical study was conducted in the emergency department and intensive care unit of a tertiary hospital. A total of 104 sepsis patients and 48 healthy adult volunteers who presented to the emergency department between April 2015 and October 2015 were included. Blood samples were collected on days 1 and 3, and BTP levels were measured using the ELISA method. Statistical analyses were performed using SPSS 22.0.

Results: BTP levels were significantly higher in sepsis patients compared to the control group (p=0.013). However, no significant difference was observed between day 1 and day 3 BTP levels (p=0.119). When categorized by sepsis severity, BTP levels did not correlate with disease severity (p>0.05). Additionally, no significant association was found between BTP levels and mortality (p=0.651).

Conclusion: BTP may serve as a potential biomarker for sepsis diagnosis, but it is not a reliable indicator of disease severity or prognosis. Further large-scale studies are needed.

Keywords: Sepsis, beta-trace protein, biomarker, prognosis, emergency department

Introduction

Sepsis is a complex condition characterized by increasing incidence, high mortality, and challenging treatment (1). Despite advances in treatment strategies and a better understanding of the molecular pathophysiology of sepsis, the growing prevalence of multidrug-resistant microorganisms, the high frequency of infections in intensive care units (ICUs), and the increasing number of immunosuppressed patients due to transplantation, chemotherapy, and radiotherapy have contributed to the rising incidence of sepsis (2).

To improve the prognosis of critically ill patients admitted to ICUs, early recognition and intervention are essential. Various advanced diagnostic methods are routinely used to facilitate early diagnosis, prevent septic shock or multiple organ failure, and enable prompt treatment, thereby increasing patient survival (3,4). The most crucial factors determining the prognosis of sepsis patients are early diagnosis and rapid treatment. Current diagnostic markers include clinical findings such as leukopenia or leukocytosis, as well as elevated C-reactive protein (CRP) and sedimentation rate. However, none of these biomarkers is specific to sepsis. Thus, there is an urgent need for novel biomarkers that can aid in the early diagnosis and treatment monitoring of sepsis, and ongoing research is focused on identifying such markers (3).

Beta-trace protein (BTP) was first identified in human cerebrospinal fluid in 1961 using immunoelectrophoresis. In 1985, a novel glycoprotein, lipocalin-type prostaglandin D2 synthase (L-PGDS),

Corresponding Author: Mehmet Yortanlı MD, Konya Numune Hospital, Clinic of Emergency Medicine, Konya, Türkiye
E-mail: myortanli@gmail.com **ORCID ID:** orcid.org/0000-0002-6744-2423
Cite this article as: Yortanlı M, Köylü R, Köylü Ö, Yortanlı BÇ, Cander B. The role of serum beta-trace protein in the diagnosis and prognosis of sepsis patients in the emergency department. Eurasian J Emerg Med. 2025;24(3): 190-5.

Received: 23.02.2025
Accepted: 01.04.2025
Epub: 05.05.2025
Published: 10.09.2025

was isolated from mouse brains. Subsequent studies confirmed that BTP and L-PGDS are identical molecules (5). BTP plays a role in the induction of, inhibition of bronchoconstriction, and platelet aggregation. Hematopoietic prostaglandin D synthase (H-PGDS) is found in mast cells, helper T-cells, Kupffer cells, dendritic cells, and microglia and is known to be more active in allergic and inflammatory responses. As a mediator of allergic and inflammatory processes, BTP also exhibits a high affinity for various lipophilic compounds such as retinoic acid, bilirubin, biliverdin, thyroid hormones, amyloid β peptides, and gangliosides, functioning as an extracellular transporter. Additionally, by binding endogenous amyloid β, it prevents in vivo amyloid accumulation. BTP functions as a dual-purpose protein due to its role as an enzyme, producing prostaglandin D2 (PGD2) and as a lipophilic ligand-binding protein. It catalyzes the conversion of prostaglandin H2, an arachidonic acid derivative, to PGD2. Despite its multiple proposed functions, the exact physiological role of BTP remains unclear (6-8).

Recent studies have demonstrated that BTP serves as a PGD2 receptor for T-helper 2 cells, inducing the chemotaxis of eosinophils, basophils, and macrophages in response to PGD2 stimulation. In addition to its chemotactic effects, PGD2 has been shown to modulate the immune response by interacting with the prostaglandin D (PGD) receptor in inflammatory cells (9). These findings suggest that BTP may play a role in the pathophysiology of sepsis and could serve as a prognostic biomarker.

Materials and Methods

This prospective, single-center study was conducted in the emergency department and ICU a tertiary care hospital. Patients who presented to the emergency department between April 2015 and October 2015 were 18 years or older, not pregnant or breastfeeding, and provided informed consent were included in the study. A total of 104 patients diagnosed with sepsis and 48 healthy adult volunteers without any known diseases were enrolled. Blood samples were collected from the patients on day 1 and day 3. The study was designed as a controlled, open-label, observational, prospective clinical study. Ethical approval was obtained from the Necmettin Erbakan University Meram Faculty of Medicine Ethic Committe on (decision number: 2015/170, date: 18.09.2015), and the study was conducted in accordance with the principles of the Declaration of Helsinki. Informed consent was obtained from all participants before their inclusion in the study.

The human BTP levels were measured using an ELISA based on biotin and double-antibody sandwich technology. The assay was designed for research purposes and utilized 96-well plates pre-coated with BTP antibodies. BTP levels were measured using YH-BIOSEARCH Human BTP ELISA kits (catalog number: YHB3523Hu, China). During the analysis, a BioTek ELx50 microplate washer (USA) was used for washing steps, and absorbance readings were obtained using a BioTek ELx800 microplate reader (USA). All parameters were measured at a wavelength of 450 nm, and absorbance values were analyzed using a calibration curve to determine the BTP concentrations in the samples.

Statistical Analysis

All data were transferred to SPSS 22.0 (IBM, USA) statistical software for analysis. The normality of data distribution was assessed using the Shapiro-Wilk test and the Kolmogorov-Smirnov test. Numerical variables were presented as mean ± standard deviation for normally distributed data, and as median (interquartile range) for non-normally distributed data, while categorical variables were expressed as percentages. Comparisons between groups were performed using the independent samples t-test for normally distributed data and the Mann-Whitney U test for non-normally distributed data. Changes within the same group over time were evaluated using the paired t-test for normally distributed data and the Wilcoxon signed-rank test for non-normally distributed data. Comparisons among three or more groups were performed using the Kruskal-Wallis test for non-normally distributed data. Categorical variables were analyzed using the chi-square (χ^2) test or Fisher's exact test, depending on the sample size and distribution. A p value of <0.05 was considered statistically significant.

Results

In our study, a statistically significant difference was observed in the mean age between the patient and control groups ($p<0.001$), whereas no significant difference was found in gender distribution ($p=0.084$). When comparing blood BTP levels, patients in the sepsis group had significantly higher BTP levels than those in the control group ($p=0.013$) (Table 1).

Among sepsis patients, a comparison of laboratory parameters between day 1 and day 3 showed statistically significant changes in white blood cell count, neutrophil count, hemoglobin, platelet count, creatinine, sodium, potassium, calcium, CRP, and procalcitonin (PCT) levels. However, no significant difference was observed in BTP levels between day 1 and day 3 ($p=0.119$) (Table 2).

When sepsis patients were categorized into groups: sepsis, severe sepsis, and septic shock, no statistically significant association was found between BTP levels and sepsis severity on day 1 and day 3 ($p=0.520$, $p=0.217$) (Table 3).

Analysis of patients' medical history revealed a borderline significant relationship between stroke and BTP levels ($p=0.052$),

whereas no significant association was found between BTP levels and coronary artery disease (CAD), hypertension, chronic obstructive pulmonary disease, diabetes mellitus, chronic kidney disease (CKD), or cancer (Table 4).

Regarding patient outcomes, 48.1% (n=50) of sepsis patients died in the hospital, while 51.9% (n=54) were discharged. However, no statistically significant relationship was observed between BTP levels and mortality (p=0.651).

Table 1. Comparison of patient and control groups

Variables	Sepsis group (n=104)	Control group (n=48)	p value*
Age, mean ± SD	70.6±15.47	53.67±9.84	**<0.001**
Gender, n (%) Male Female	 59 (56.7%) 45 (43.3%)	 20 (41.7%) 28 (58.3%)	 0.084
BTP, pg/mL, median (IQR)	72.6 (45.05)	66.07 (23.92)	**0.013**

*Statistically significant p values are highlighted in bold. BTP: Beta-trace protein, IQR: Interquartile range, SD: Standard deviation

Table 2. Comparison of day 1 and day 3 laboratory parameters in sepsis patients

Variable	Day 1	Day 3	p value*
Apache II score, mean ± SD	18.6±7.249	18.96±8.235	0.378
GCS, median (IQR)	13 (5)	14 (4)	0.605
WBC, 10^3/µL, median (IQR)	13.95 (12.11)	11.64 (8.38)	**0.020**
Neutrophil, 10^3/µL, median (IQR)	11.45 (9.79)	9.69 (8.3)	**0.009**
Lymphocyte, 10^3/µL, median (IQR)	0.84 (0.91)	1.12 (0.81)	0.053
Hemoglobin, g/dL, mean ± SD	12.34±2.31	11.27±1.9	**<0.001**
Platelet, 10^3/µL, median (IQR)	212 (169)	183 (173)	**0.006**
Creatinine, mg/dL, median (IQR)	1.81 (2.08)	1.13 (1.42)	**<0.001**
pH, mean ± SD	7.39±0.11	7.39±0.09	0.890
pCO_2, mmHg, mean ± SD	34.96±10.19	36.36±9.19	0.151
HCO_3, mmol/L, mean ± SD	20.32±4.19	20.76±4.64	0.919
Lactate, mmol/L, median (IQR)	1.1 (1.05)	1.1 (1.07)	0.676
Sodium, mEq/L, mean ± SD	135.84±9.08	137.88±6.49	**0.011**
Potassium, mEq/L, mean ± SD	4.48±1.02	4.02±0.89	**<0.001**
Calcium, mEq/L, mean ± SD	8.01±0.86	7.59±0.83	**<0.001**
Albumin, g/dL, mean ± SD	2.67±0.6	2.48±0.41	0.057
Sedimentation rate, mm/h, median (IQR)	54 (49)	62 (51)	0.431
CRP, mg/L, median (IQR)	139 (90)	116 (105.6)	**0.001**
Procalcitonin, mg/dL, median (IQR)	6.75 (29.77)	4.79 (16.64)	**<0.001**
BTP, pg/mL, median (IQR)	72.6 (45.05)	69.1 (39.34)	0.119
Mechanical ventilation, n (%)	30 (%28.8)	21 (%25.3)	0.625
Vasopressor use, n (%)	53 (%51)	32 (%38.6)	0.118

*Statistically significant p values are highlighted in bold. APACHE: Acute physiology and chronic health evaluation, GCS: Glasgow coma scale, WBC: White blood cell, CRP: C-reactive protein, BTP: Beta-trace protein, IQR: Interquartile range, SD: Standard deviation

Table 3. Comparison of BTP levels on day 1 by sepsis severity

Sepsis severity	N (%)	BTP, pg/mL, median (IQR)	p value
Sepsis	53 (%51)	78.96 (36.27)	0.520*
Severe sepsis	27 (%26)	59.26 (51.31)	
Septic shock	24 (%23)	56.92 (89.62)	

*p values were calculated using the Kruskal-Wallis test. BTP: Beta-trace protein, IQR: Interquartile range

Table 4. Comparison of BTP levels based on patient history			
Diagnosis	**n (%)***	**BTP, pg/mL, median (IQR)***	**p value**
CAD	20 (19.2%) / 84 (80.8%)	83.61 (66.83) / 68.87 (42.27)	0.466
Hypertension	44 (42.3%) / 60 (57.7%)	73.93 (99.85) / 67.32 (42.74)	0.177
Stroke	25 (24%) / 79 (76%)	81.52 (66.11) / 67.24 (40.22)	0.052
COPD	32 (30.8%) / 72 (69.2%)	71.63 (43.35) / 70.35 (50.76)	0.776
DM	29 (27.9%) / 75 (72.1%)	61.39 (52.56) / 70.81 (50.10)	0.821
CKD	30 (28.8%) / 74 (71.2%)	59.26 (45.22) / 74.50 (51.18)	0.249
Cancer	9 (8.7%) / 95 (91.3%)	67.86 (30.00) / 73.06 (53.21)	0.447

*Patients with/without the disease, BTP: Beta-trace protein, CAD: Coronary artery disease, COPD: Chronic obstructive pulmonary disease, DM: Diabetes mellitus, CKD Chronic kidney disease, IQR: Interquartile range

Discussion

Sepsis is a critical syndrome with high mortality rates that arises from a dysregulated host response to infection (10). Early intervention in the diagnosis of sepsis is of vital importance; however, the sensitivity and specificity of existing biomarkers are still insufficient (11). In our study, BTP was found to be significantly elevated in septic patients, suggesting that this biomarker may play a role in the diagnosis of sepsis.

The pathophysiological mechanisms of sepsis remain complex and arenot fully understood. Traditionally, sepsis has been defined as a systemic inflammatory response to infection, yet in some cases, the source of infection cannot be identified (12). PGD, prostaglandin D synthase (PGDS), cyclooxygenase-1 and cyclooxygenase-2 play crucial roles in eicosanoid metabolism during infection. Since BTP, also known as L-PGDS, is linked to infection responses, it is hypothesized that it may play a role in the immune and inflammatory pathways of sepsis. Moreover, previous studies have demonstrated its involvement in thrombotic events, neurodegenerative diseases, kidney function, and inflammatory processes (5-8,13).

The current literature indicates that BTP plays a significant role in inflammatory processes and has important effects on vascular endothelium (14). Studies have reported that BTP actively participates in inflammatory processes, exhibiting pro-angiogenic and anti-inflammatory effects. BTP is produced by endothelial cells and contributes to the modulation of inflammatory responses. Similarly, the activation of PGD2 receptors has been shown to trigger strong inflammatory reactions in macrophages (5,15). Our findings support this, as BTP levels were significantly higher in sepsis patients compared to healthy controls. This suggests that BTP may contribute to inflammatory responses and could be useful as a diagnostic biomarker for sepsis.

The role of biomarkers in the diagnosis of sepsis has been investigated for a long time. In the existing literature, biomarkers such as PCT and CRP have been frequently studied for sepsis diagnosis and prognosis. PCT has been reported to have better diagnostic value in the early diagnosis of sepsis compared to CRP. However, it has been shown that these biomarkers alone do not possess sufficient sensitivity and specificity (11,16).

The molecular and genetic differences among septic patients highlight the necessity of developing biomarker-based diagnostic systems. In recent years, it has been suggested that classifying septic patients into more specific subgroups through phenotyping and endotyping could improve treatment outcomes (17). In our study, the significantly elevated BTP levels in septic patients suggest that novel biomarkers, supported by molecular and genetic research, could contribute to the diagnostic process.

Moreover, considering the critical role of endothelial cells in the pathophysiology of sepsis, the role of BTP in vascular inflammation should be further investigated. Microvascular dysfunction and coagulation activation in sepsis are among the key factors that exacerbate disease severity (18). A study conducted by Bruegel et al. (19) demonstrated that arachidonic acid metabolism undergoes significant changes in sepsis and that these alterations may be associated with disease severity and clinical prognosis. Similarly, in the study by Kinoshita et al. (20), L-PGDS levels were found to be related to endothelial damage and identified as a potential diagnostic biomarker in preeclampsia. Given that similar mechanisms may be involved in septic patients, the contribution of BTP to vascular inflammation processes should be investigated in greater detail.

However, White et al. (6) demonstrated that BTP levels fluctuate during acute inflammatory conditions such as sepsis, limiting their usefulness in disease monitoring. Similarly, in our study, no significant difference was observed in BTP levels between day 1 and day 3, and BTP was not associated with mortality. This suggests that BTP may not be a reliable marker for disease progression or prognosis in sepsis.

Ahmad et al. (13) found that eicosanoid metabolites increase with sepsis severity but exhibit variability in prolonged inflammatory processes. In line with this, our study demonstrated that BTP levels were not correlated with disease severity, further suggesting that while BTP may assist in the diagnosis of sepsis, it may not be a reliable indicator.

Regarding kidney function, studies have shown that BTP is not cleared by hemodialysis and may serve as a biomarker for kidney injury in CKD patients (21,22). In addition, increased H-PGDS expression has been observed in skeletal, and cardiac muscle cells, although no significant association between BTP levels and CAD has been identified (7,23). Furthermore, BTP has been linked to neurodegenerative processes and has been suggested to have a protective role against cerebral ischemia (24). In our study, no significant relationship was found between BTP levels and CKD or CAD, whereas a borderline significant association was observed between BTP and stroke.

Study Limitations

Our study has several limitations. The sample size is relatively small, and the study was conducted at a single center. Larger, multicenter studies are needed to validate these findings. Since our control group consisted entirely of healthy individuals, it was not possible to compare the results with critically ill non-septic patients. To better understand the specificity of sepsis-related biomarkers, future studies should include non-septic critically ill patients. The case-control ratio in our study is unbalanced due to difficulties in recruiting healthy volunteers. Larger-scale studies are required to achieve more robust statistical analyses and generalizable results. The average age of individuals in the control group is lower than that of septic patients. There is no clear consensus in the literature regarding how BTP levels change with age. However, to better understand the impact of age on the results, future studies should include age-matched control groups. The data in our study were collected in 2015, and the effects of updated sepsis management protocols and new treatment approaches on BTP levels have not been evaluated. Therefore, further studies are needed to determine the diagnostic and prognostic value of BTP in current clinical practice.

Conclusion

Our study demonstrated that BTP levels are elevated in sepsis patients compared to healthy controls, suggesting their potential as a diagnostic biomarker. However, BTP was not associated with disease severity or mortality, limiting its use as a prognostic marker. Further large-scale, multicenter studies are required to better understand the diagnostic and clinical utility of BTP in sepsis.

Ethics

Ethics Committee Approval: Ethical approval was obtained from the Local Ethics Committee on (decision number: 2015/170, date: 18.09.2015), and the study was conducted in accordance with the principles of the Declaration of Helsinki.

Informed Consent: Informed consent was obtained from all participants before their inclusion in the study.

Footnotes

Author Contributions

Surgical and Medical Practices: M.Y, R.K., Concept: M.Y., R.K., B.Ç.Y., Design: M.Y.,B.Ç.Y.,B.C., Data Collection or Processing: M.Y.,Ö.K., B.Ç.Y., Analysis or Interpretation: M.Y., R.K., Ö.K., B.Ç.Y.,B.C., Literature Search: M.Y., R.K., Ö.K., B.Ç.Y.,B.C., Writing: M.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. Srzić I, Nesek Adam V, Tunjić Pejak D. Sepsis definition: what's new in the treatment guidelines. Acta Clin Croat. 2022;61(Suppl 1):67-72.
2. Chiu C, Legrand M. Epidemiology of sepsis and septic shock. Curr Opin Anaesthesiol. 2021;34:71-6.
3. Faix JD. Biomarkers of sepsis. Crit Rev Clin Lab Sci. 2013;50:23-36.
4. Oczkowski S, Alshamsi F, Belley-Cote E, Centofanti JE, Hylander Møller M, Nunnaly ME, et al. Surviving Sepsis Campaign Guidelines 2021: highlights for the practicing clinician. Pol Arch Intern Med. 2022;132:16290.
5. Urade Y, Fujimoto N, Hayaishi O. Purification and characterization of rat brain prostaglandin D synthetase. J Biol Chem. 1985;260:12410-5.
6. White CA, Ghazan-Shahi S, Adams MA. β-Trace protein: a marker of GFR and other biological pathways. Am J Kidney Dis. 2015;65:131-46.
7. Sert ET, Akilli N, Köylü R, Cander B, Kokulu K, Köylü Ö. The effect of beta-trace protein on diagnosis and prognosis in patients with acute coronary syndrome. Cureus. 2020;12:e7135.
8. Almela MT, Navarro-Zaragoza J, Laorden ML, Sánchez-Celemín F, Almela P. Cut-off value for β-trace protein (β-TP) as a rapid diagnostic of cerebrospinal fluid (CSF) leak detection. Laryngoscope Investig Otolaryngol. 2023;8:1233-9.
9. Hokari R, Nagata N, Kurihara C, Watanabe C, Komoto S, Okada Y, et al. Increased expression and cellular localization of lipocalin-type prostaglandin D synthase in helicobacter pylori-induced gastritis. J Pathol. 2009;219:417-26.
10. Cohen M, Banerjee D. Biomarkers in sepsis: a current review of new technologies. J Intensive Care Med. 2024;39:399-405.
11. Barber G, Tanic J, Leligdowicz A. Circulating protein and lipid markers of early sepsis diagnosis and prognosis: a scoping review. Curr Opin Lipidol. 2023;34:70-81.
12. 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.

13. Ahmad NS, Tan TL, Arifin KT, Ngah WZW, Yusof YAM. High sPLA2-IIA level is associated with eicosanoid metabolism in patients with bacterial sepsis syndrome. PLoS One. 2020;15:e0230285.
14. Li J, Li C, Subedi U, Subedi P, Panchatcharam M, Sun H. The role of endothelial L-PGDS in the pro-angiogenic and anti-inflammatory effects of low-dose alcohol consumption. Cells. 2024;13:2007.
15. Jandl K, Stacher E, Bálint Z, Sturm EM, Maric J, Peinhaupt M, et al. Activated prostaglandin D2 receptors on macrophages enhance neutrophil recruitment into the lung. J Allergy Clin Immunol. 2016;137:833-43.
16. Rintala EM, Aittoniemi J, Laine S, Nevalainen TJ, Nikoskelainen J. Early identification of bacteremia by biochemical markers of systemic inflammation. Scand J Clin Lab Invest. 2001;61:523-30
17. Scherger SJ, Kalil AC. Sepsis phenotypes, subphenotypes, and endotypes: are they ready for bedside care? Curr Opin Crit Care. 2024;30:406-13.
18. Dolmatova EV, Wang K, Mandavilli R, Griendling KK. The effects of sepsis on endothelium and clinical implications. Cardiovasc Res. 2021;117:60-73.
19. Bruegel M, Ludwig U, Kleinhempel A, Petros S, Kortz L, Ceglarek U, et al. Sepsis-associated changes of the arachidonic acid metabolism and their diagnostic potential in septic patients. Crit Care Med. 2012;40:1478-86.
20. Kinoshita K, Takeda J, Matsuoka K, Takeda S, Eguchi Y, Oda H, et al. Expression of lipocalin-type prostaglandin D synthase in preeclampsia patients: a novel marker for preeclampsia. Hypertens Res Pregnancy. 2014;2:72–7.
21. Geethanjali R, Ganesh M, Vinod A, Ramprasad E, Jayaprakash V. Correlation of beta trace protein levels with serum creatinine-based estimated glomerular filtration rate equations in chronic kidney disease. G Ital Nefrol. 2023;40:2023-vol6.
22. Iversen E, Walls AB, Petersen A, Jensen PS, Kallemose T, Andersen A, et al. Estimated glomerular filtration rate based on creatinine, cystatin C, β-trace protein and β2 microglobulin in patients undergoing nontraumatic lower extremity amputation. Br J Clin Pharmacol. 2023;89:1789-98.
23. Hamamura K, Yoshida Y, Oyama K, Li J, Kawano S, Inoue K, et al. Hematopoietic prostaglandin D synthase is increased in mast cells and pericytes in autopsy myocardial specimens from patients with duchenne muscular dystrophy. Int J Mol Sci. 2024;25:1846.
24. Saleem S, Shah ZA, Urade Y, Doré S. Lipocalin-prostaglandin D synthase is a critical beneficial factor in transient and permanent focal cerebral ischemia. Neuroscience. 2009;160:248-54.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.14237
Eurasian J Emerg Med. 2025;24(3): 196-200

Comparison of Surgical and Non-Surgical Approaches on Mortality in Hip Fracture Patients

Doğuhan Bitlisli, Muhammed İkbal Şaşmaz, Bülent Demir

Manisa Celal Bayar University Faculty of Medicine, Department of Emergency Medicine, Manisa, Türkiye

Abstract

Aim: Hip fractures are a prevalent and severe health issue, especially in the elderly population. The incidence of hip fractures is expected to increase globally, posing a significant public health challenge. Surgical intervention is traditionally the standard treatment, though the effectiveness of surgical versus non-surgical management in reducing mortality remains unclear. This study aimed to evaluate the effects of surgical and non-surgical approaches on one-month and one-year mortality rates in patients with hip fractures.

Materials and Methods: A prospective, single-center study was conducted at Manisa Celal Bayar University Faculty of Medicine Hospital, including 113 patients diagnosed with hip fractures. Patients were classified based on whether they received surgical intervention or not.

Results: The mean age of patients was 67.91±19.95 years, with 63 (55.8%) women. Among the patients, 94 (83.2%) underwent surgery, and 19 (16.8%) did not. The one-month and the one-year mortality rates were 13.3%, and 22.1%, respectively. For surgical patients, one-month and one-year mortality rates were 8.5% and 16%, respectively. In contrast, non-surgical patients had higher mortality rates, with 36.8% one-month and 52.6% one-year mortality. The difference in mortality between surgical and non-surgical patients was statistically significant ($p<0.001$ for both).

Conclusion: Surgical treatment for hip fractures significantly reduces both one-month and one-year mortality compared to non-surgical management. While surgery is beneficial, the decision should be individualized, considering patient comorbidities and overall health. For frail, high-risk patients, non-surgical approaches may be more appropriate. This study emphasizes the importance of timely surgical intervention to improve survival outcomes in hip fracture patients.

Keywords: Hip fractures, surgical approach, non-surgical approach

Introduction

Hip fractures are a significant global health problem that especially affects older adults in their eighties (1). As life expectancy increases world-wide, the incidence of hip fractures is expected to increase in the coming years. The likelihood of suffering a hip fracture increases with age; over the age of 50, the incidence doubles with each passing decade. The increase in the incidence of the disease, along with the aging of the population, has reached widespread, even epidemic proportions, creating a major public health problem. This situation has been called a "massive public health crisis", consuming a large portion of healthcare resources along with a huge burden of suffering for patients and their family members. It is estimated that the incidence of hip fractures world-wide will increase to 6.3 million per year by 2050 (2). Hip fractures have a higher mortality rate in older adults, with a 30-day cumulative mortality rate of 14%. This rate can reach 30% in the first year postoperatively (3,4). Additionally, serious postoperative complications occur in approximately 20% of patients (5,6). Hip fractures have historically been treated with surgery. However, mortality and morbidity outcomes after surgery are poor, especially in elderly patients with certain risk factors (7,8). These risk factors include older age, American Society of Anesthesiologists score greater than 3 or 4, low mobility, and cognitive impairment. In the context of palliative care, non-surgical treatments are increasingly being offered as an alternative to the operative approach in these

Corresponding Author: Bülent Demir MD, Manisa Celal Bayar University Faculty of Medicine, Department of Emergency Medicine, Manisa, Türkiye
E-mail: dr_bulent@yahoo.com **ORCID ID:** orcid.org/0000-0001-8305-2882
Cite this article as: Bitlisli D, Şaşmaz Mİ, Demir B. Comparison of surgical and non-surgical approaches on mortality in hip fracture patients. Eurasian J Emerg Med. 2025;24(3): 196-200.

Received: 25.03.2025
Accepted: 22.04.2025
Epub: 19.06.2025
Published: 10.09.2025

frail geriatric hip fracture patients with very limited life expectancy (9). The incidence of hip fracture is approximately three times higher in women than in men (10). A number of other risk factors, such as a history of previous hip fracture, poor diet, white race, alcohol and caffeine consumption, physical inactivity, tall stature, and visual impairment, play a role in the distribution of specific risk ratios in different subpopulations. A typical fracture occurs in an elderly patient with weakened and osteoporotic bones (90% of hip fractures occur in patients over 65 years of age), following a simple and relatively minor fall (e.g., falling from a bed or chair or from a standing position) (10,11). The management of such fractures depends on the characteristics of the fracture (location, displacement, comminution, and bone quality) and the patient's age and condition (12). However, there is general agreement that hip fractures should be treated surgically, either with internal fixation or endoprosthetic replacement. However, especially in bedridden and nursing patients, no surgical intervention is justified. An increased risk of death after hip fracture surgery has been previously recognized (13-15). A recent large study found that the acute inpatient mortality rate was approximately 3.1% (16). Estimates of mortality in elderly patients 1 year after hip fracture vary between 14% and 36% in different studies (17). Hip fracture mortality has been found to be higher in men than in women (10). Other risk factors for death after hip fracture include severe systemic disease, psychiatric illness, postoperative complications, and residence in a nursing home (18). The aim of this study was to evaluate the effects of surgical repair and non-surgical approaches on one-month and one-year mortality in hip fracture cases.

Materials and Methods

Study Design and Population

This prospective, single-center study was carried out at the Manisa Celal Bayar University Faculty of Medicine Hospital, a tertiary care academic medical center. The research received ethical approval from the Health Sciences Ethics Committee at Manisa Celal Bayar University Faculty of Medicine (desicion number: 20478486/1969, date: 31.08.2023). Informed written consent was obtained from the patients or their immediate family members who volunteered to participate in the study. Patients diagnosed with hip fractures who agreed to participate in the study were included. Unstable patients and those under the age of 18 were excluded from the study. The patients' demographic characteristics, mechanisms and timing of the injury, associated injuries, and fracture types were recorded in the study form. The Injury Severity score (ISS), and the 1-month and 1-year mortality rates were compared between patients based on whether they underwent surgery or not.

Statistical Analysis

Data analyses were conducted using the Statistical Package for the Social Sciences Software (version 26.0). Descriptive statistics, including frequency, percentage, mean, standard deviation (SD), median, minimum, and maximum values, were used to summarize the data. The normality of the data was assessed using the Kolmogorov-Smirnov test. In univariate analysis, continuous variables with a normal distribution were presented as mean ± SD and compared using the t-test. Categorical variables were analyzed using the Pearson chi-square test, while Fisher's exact test was applied when the frequency of categories was less than five. A p value of <0.05 was considered statistically significant.

Results

A total of 113 patients were included in the study, with a mean age of 67.91±19.95, and 63 (55.8%) were women. Upon examining the comorbidities of the patients, hypertension was the most common (58.4%), followed by diabetes mellitus (29.2%). The majority of patients presented to our emergency department due to falls from the ground level (77.9%) and between the hours of 23:00-07:00, (54%). Trochanteric fractures were identified in 62 (54.9%) patients, while femoral neck fractures were identified in 51 (45.1%) patients. A total of 94 (83.2%) patients underwent surgery. The one-month mortality rate was 13.3%, while the one-year mortality rate was 22.1% (Table 1). In the majority of patients (n=94) an isolated hip fracture was detected, while 19 patients (16.8%) had associated injuries. When the associated trauma regions were examined, the most common injury was to the thorax (n=11), followed by head and neck injuries (Table 2). The mean age of 94 (83.2%) patients who underwent surgery and 19 (16.8%) patients who did not undergo surgery was 65.82±20.90 and 78.26±9.25, respectively ($p<0.001$), while the mean ISS score was 10.40±4.24 and 12.32±6.16, respectively ($p=0.210$). When evaluating the one-month and one-year mortalities of patients who underwent surgery and those who did not, the one-month mortality for patients who underwent surgery was 8.5%, and the one-year mortality was 16%, while the one-month mortality for patients who did not undergo surgery was 36.8%, and the one-year mortality was 52.6%. In patients who underwent surgery, both the one-month and one-year mortality rates were significantly lower ($p<0.001$, $p<0.001$, respectively) (Table 3). When evaluating the one-year mortality rates according to the type of fracture, mortality was observed in 17 out of 62 patients (27.4%) with trochanteric fractures, and in 8 out of 51 patients (15.7%) with femoral neck fractures ($p=0.135$).

Table 1. Demographic and clinical data of the patients		
	Mean ± SD	
Age	67.91±19.95	
ISS score	10.72±4.64	
		n (%)
Gender	Female Male	63 (55.8%) 50 (44.2%)
Comorbidities	Hypertension Diabetes mellitus Ischemic heart disease Stroke Dementia Parkinson's disease	66 (58.4%) 33 (29.2%) 16 (14.2%) 9 (8%) 9 (8%) 6 (5.3%)
Mechanism of injury	Fall from ground level Fall from a height In-vehicle traffic accident Motorcycle accident	88 (77.9%) 10 (8.8%) 8 (7.1%) 7 (6.2%)
Time of injury	23:00-07:00 07:00-15:00 15:00-23:00	61 (54%) 23 (20.3%) 29 (25.7%)
Type of fracture	Trochanteric fracture Femur neck fracture	62 (54.9%) 51 (45.1%)
Surgery	Yes No	94 (83.2%) 19 (16.8%)
One-month mortality	Yes No	15 (13.3%) 98 (86.7%)
One-year mortality	Yes No	25 (22.1%) 88 (77.9%)
ISS: The injury severity score, SD: Standard deviation		

Table 2. Associated injuries	
Trauma region	**n (%)**
Head and neck	9 (8%)
Thorax	11 (9.7%)
Abdomen	5 (4.4%)
Pelvic	5 (4.4%)
Vertebra	4 (3.5%)
Upper extremity	5 (4.4%)
Lower extremity	4 (3.5%)

Discussion

Hip fractures, particularly in the aging population, represent a significant challenge for healthcare systems globally. As the global life expectancy continues to rise, the incidence of hip fractures is expected to increase, further complicating the management of these patients. Traditionally, surgical intervention has been considered the standard treatment for hip fractures. However, there is growing recognition that treatment decisions should be individualized based on patient factors, particularly in regard to their risk of mortality (19). The primary aim of this study was to assess the impact of surgical versus non-surgical treatment on both short-term (one-month) and long-term (one-year) mortality outcomes in patients with hip fractures. The findings suggest that surgical treatment is associated with a substantial reduction in mortality at both one month and one year. Specifically, the one-month mortality rate for patients who underwent surgery was 8.5%, in contrast to 36.8% for non-surgical patients. Similarly, the one-year mortality rate was 16% for surgical patients compared to 52.6% for non-surgical patients. These results are in line with previous studies, which have shown a significant reduction in mortality following surgical intervention, particularly among patients who are not terminally ill or excessively frail to undergo surgery (20-22). However, it is essential to recognize that not all patients with hip fractures are candidates for surgery. Decisions regarding surgical intervention should be made on a case-by-case basis, taking into account the patient's overall health status, comorbid conditions, and life expectancy. Literature suggests that frail patients with severe comorbidities, such as advanced dementia or significant cardiovascular disease, are at a heightened risk of complications post-surgery, which may contribute to poorer outcomes (23,24). In these cases, non-surgical management focused on pain relief and comfort, as well as palliative care, may be more appropriate to optimize quality of life. The mortality rates observed in our study align closely with those seen in other researches, underscoring the significant impact of hip fractures on the elderly population (20). A systematic epidemiological review indicates that the

Table 3. Comparison of patients who underwent surgery and those who did not					
Surgery	**Yes (n=94)**		**No (n=19)**		**p value**
Age (mean ± SD)	65.82±20.90		78.26±9.25		<0.001
ISS score (mean ± SD)	10.40±4.24		12.32±6.16		0.210
1-month mortality (n, %)	**Pozitive**	**Negative**	**Pozitive**	**Negative**	<0.001
	8 (8.5%)	86 (91.5%)	7 (36.8%)	12 (63.2%)	
1-year mortality (n, %)	**Pozitive**	**Negative**	**Pozitive**	**Negative**	<0.001
	15 (16%)	79 (84%)	10 (52.6%)	9 (47.4%)	
ISS: The injury severity score, SD: Standard deviation					

one-year mortality rate following surgically treated hip fractures ranges from 8.4% to 36%, depending on various risk factors, including age, comorbidities, and the presence of postoperative complications (25). In our study, surgical patients had a lower one-year mortality rate (16%) compared to non-surgical patients (52.6%), suggesting that surgical intervention offers a substantial survival benefit, particularly when performed in a timely manner. Our study also observed that patients with trochanteric fractures experienced a higher mortality rate (27.4%) compared to those with femoral neck fractures (15.7%). While this difference did not reach statistical significance (p=0.135), it raises the question of whether fracture type influences long-term survival outcomes. Femoral neck fractures are often more complex and associated with a higher risk of complications, such as non-union and infection, which could contribute to higher mortality rates (26,27). Nonetheless, the findings from our study suggest that the type of fracture may play a role in mortality, but further research is needed to better understand the underlying mechanisms and potential treatment implications. Comorbid conditions significantly influence the prognosis of hip fracture patients (28). In our study, hypertension was the most prevalent comorbidity (58.4%), followed by diabetes mellitus (29.2%). These chronic conditions are known to complicate the treatment and recovery process, especially in older adults. Prior research has shown that comorbidities, such as cardiovascular disease, diabetes, and respiratory illnesses, are associated with higher mortality rates following hip fractures (28,29). Moreover, patients who underwent surgery in our study were significantly younger than those who did not (65.82 years vs. 78.26 years), which may explain the better outcomes observed in the surgical group. Our findings also reflect the common mechanism of injury for hip fractures in the elderly, with 77.9% of patients sustaining their fractures from ground-level falls. Osteoporosis, which increases fracture risk, along with the physical frailty of older adults, contributes to the high incidence of hip fractures (30,31). Preventive measures, such as osteoporosis management, fall prevention programs, and modifications to home environments, may help reduce the frequency of hip fractures and their associated complications in the future.

Study Limitations

While the study provides valuable insights, it has several limitations. The single-center design and relatively small sample size (113 patients) may limit the generalizability of our findings. Larger, multicenter studies are needed to validate these results and further explore the relationship between surgical interventions, fracture type, and mortality outcomes. Additionally, we did not assess functional outcomes or quality of life after treatment, which are important factors when evaluating the overall success of treatment strategies for hip fractures. Future studies should incorporate these endpoints to provide a more comprehensive view of the effectiveness of surgical versus non-surgical approaches.

Conclusion

In conclusion, our study reinforces the significant role of surgical intervention in improving survival outcomes for patients with hip fractures. Despite the potential benefits of surgery, it is essential to evaluate each patient individually, considering their overall health, comorbidities, and functional status. For frail patients with limited life expectancy, non-surgical management, including palliative care, may be more appropriate. Ultimately, this study underscores the need for timely surgical intervention, combined with careful postoperative management, to optimize outcomes for hip fracture patients. Future research should focus on refining treatment strategies and identifying patient subgroups who would benefit most from either surgical or non-surgical approaches.

Ethics

Ethics Committee Approval: The research received ethical approval from the Health Sciences Ethics Committee at Manisa Celal Bayar University Faculty of Medicine (desicion number: 20478486/1969, date: 31.08.2023).

Informed Consent: Informed written consent was obtained from the patients or their immediate family members who volunteered to participate in the study.

Footnotes

Author Contributions

Concept: D.B., M.İ.Ş., B.D., Design: M.İ.Ş., B.D., Data Collection or Processing: D.B., Analysis or Interpretation: M.İ.Ş., Literature Search: D.B., B.D., Writing: M.İ.Ş., B.D.

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

Financial Disclosure: This study was supported by Manisa Celal Bayar University (project number: 20231201).

References

1. U.S. Department of Health and Human Services. Healthy People 2010: Understanding and Improving Health. 2000.
2. Cheng SY, Levy AR, Lefaivre KA, Guy P, Kuramoto L, Sobolev B. Geographic trends in incidence of hip fractures: a comprehensive literature review. Osteoporos Int. 2011;22:2575-86.
3. Van Waesberghe J, Stevanovic A, Rossaint R, Coburn M. General vs. neuraxial anaesthesia in hip fracture patients: a systematic review and meta-analysis. BMC Anesthesiol. 2017;17:87.

4. Brauer CA, Coca-Perraillon M, Cutler DM, Rosen AB. Incidence and mortality of hip fractures in the United States. JAMA. 2009;302:1573-9.
5. Seitz DP, Gill SS, Bell CM, Austin PC, Gruneir A, Anderson GM et al. Postoperative medical complications associated with anesthesia in older adults with dementia. J Am Geriatr Soc. 2014;62:2102-9.
6. Desai V, Chan PH, Prentice HA, Zohman GL, Diekmann GR, Maletis GB, et al. Is anesthesia technique associated with a higher risk of mortality or complications within 90 days of surgery for geriatric patients with hip fractures? Clin Orthop Relat Res. 2018;476:1178-88.
7. Nkanang B, Parker M, Parker E, Griffiths R. Perioperative mortality for patients with a hip fracture. Injury. 2017;48:2180-3.
8. Cram P, Yan L, Bohm E, Kuzyk P, Lix LM, Morin SN, et al. Trends in operative and nonoperative hip fracture management 1990-2014: a longitudinal analysis of manitoba administrative data. J Am Geriatr Soc. 2017;65:27-34.
9. Ko FC, Morrison RS. Hip fracture: a trigger for palliative care in vulnerable older adults. JAMA Intern Med. 2014;174:1281-2.
10. Dennison E, Cooper C. Epidemiology of osteoporotic fractures. Horm Res. 2000;54 (Suppl 1):58-63.
11. Walter LC, Lui LY, Eng C, Covinsky KE. Risk of hip fracture in disabled community-living older adults. J Am Geriatr Soc. 2003;51:50-5.
12. Koval KJ, Zuckerman JD. Hip fractures: I. Overview and evaluation and treatment of femoral-neck fractures. J Am Acad Orthop Surg. 1994;2:141-9.
13. Bredahl C, Nyholm B, Hindsholm KB, Mortensen JS, Olesen AS. Mortality after hip fracture: results of operation within 12 h of admission. Injury. 1992;23:83-6.
14. Forster MC, Calthorpe D. Mortality following surgery for proximal femoral fractures in centenarians. Injury. 2000;31:537-9.
15. Hamlet WP, Lieberman JR, Freedman EL, Dorey FJ, Fletcher A, Johnson EE. Influence of health status and the timing of surgery on mortality in hip fracture patients. Am J Orthop (Belle Mead NJ). 1997;26:621-7.
16. Bhattacharyya T, Iorio R, Healy WL. Rate of and risk factors for acute inpatient mortality after orthopaedic surgery. J Bone Joint Surg Am. 2002;84:562-72.
17. Cumming RG, Nevitt MC, Cummings SR. Epidemiology of hip fractures. Epidemiol Rev. 1997;19:244-57.
18. Nightingale S, Holmes J, Mason J, House A. Psychiatric illness and mortality after hip fracture. Lancet. 2001;357:1264-5.
19. Roberts KC, Brox WT, Jevsevar DS, Sevarino K. Management of hip fractures in the elderly. J Am Acad Orthop Surg. 2015;23:131-7.
20. Mattisson L, Bojan A, Enocson A. Epidemiology, treatment and mortality of trochanteric and subtrochanteric hip fractures: data from the Swedish fracture register. BMC Musculoskelet Disord. 2018;19:369.
21. Frenkel Rutenberg T, Assaly A, Vitenberg M, Shemesh S, Burg A, Haviv B, et al. Outcome of non-surgical treatment of proximal femur fractures in the fragile elderly population. Injury. 2019;50:1347-52.
22. Saul D, Riekenberg J, Ammon JC, Hoffmann DB, Sehmisch S. Hip fractures: therapy, timing, and complication spectrum. Orthop Surg. 2019;11:994-1002.
23. Koso RE, Sheets C, Richardson WJ, Galanos AN. Hip fracture in the elderly patients: a sentinel event. Am J Hosp Palliat Care. 2018;35:612-9.
24. Castronuovo E, Pezzotti P, Franzo A, Di Lallo D, Guasticchi G. Early and late mortality in elderly patients after hip fracture: a cohort study using administrative health databases in the Lazio region, Italy. BMC Geriatr. 2011;5;11:37.
25. Ishizawa R, Nakanishi N, Keibun L, Sonoo T, Nakamura K, Goto T. Characteristics of patients with hip fractures and comorbid fall-related injuries in the emergency department. Acute Med Surg. 2022;27;9:e805.
26. Cornwall R, Gilbert MS, Koval KJ, Strauss E, Siu AL. Functional outcomes and mortality vary among different types of hip fractures: a function of patient characteristics. Clin Orthop Relat Res. 2004:64-71.
27. Karagiannis A, Papakitsou E, Dretakis K, Galanos A, Megas P, Lambiris E, et al. Mortality rates of patients with a hip fracture in a southwestern district of Greece: ten-year follow-up with reference to the type of fracture. Calcif Tissue Int. 2006;78:72-7.
28. Roche JJ, Wenn RT, Sahota O, Moran CG. Effect of comorbidities and postoperative complications on mortality after hip fracture in elderly people: prospective observational cohort study. BMJ. 2005;10;331:1374.
29. Dailiana Z, Papakostidou I, Varitimidis S, Michalitsis S, Veloni A, Malizos K. Surgical treatment of hip fractures: factors influencing mortality. Hippokratia. 2013;17:252-7.
30. Marks R, Allegrante JP, Ronald MacKenzie C, Lane JM. Hip fractures among the elderly: causes, consequences and control. Ageing Res Rev. 2003;2:57-93.
31. Sarvi MN. Hip fracture: anatomy, causes, and consequences. In: Total hip replacement-an overview. IntechOpen. 2018.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.60320
Eurasian J Emerg Med. 2025;24(3): 201-7

Large Language Models in Generating Differential Diagnoses in the Emergency Department: A Comparative Study of ChatGPT, Copilot, and Emergency Physician

Banu Arslan[1], Büşra Erdem[2], Merve Osoydan Satıcı[2], Çağatay Nuhoğlu[2], Hasan Yasin Soylu[1]

[1]University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye
[2]University of Health Sciences Türkiye, Şişli Hamidiye Etfal Training and Research Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye

Abstract

Aim: Accurate diagnosis in emergency departments relies heavily on clinical decision-making, yet cognitive errors contribute to a significant proportion of diagnostic mistakes. Since their launch, Generative Pre-trained Transformer-4 (GPT-4) based large language models (LLMs) have been reshaping healthcare, offering improvements in diagnostic accuracy, treatment planning, and patient care. This study evaluates the performance of these tools in generating primary and differential diagnoses compared to an experienced emergency medicine (EM) physician.

Materials and Methods: We conducted a retrospective cross-sectional study using 468 real-world clinical vignettes from non-trauma adult patients. GPT-4-based Chat Generative Pre-trained Transformer (ChatGPT) and Copilot were tasked with generating five differential diagnoses for each vignette. Their accuracy was compared to the diagnoses provided by EM physicians, using discharge diagnoses as the reference. Statistical analysis included descriptive statistics and Cohen's kappa to assess agreement.

Results: ChatGPT and Copilot demonstrated high accuracy, with correct diagnoses in the top three positions in 91.9% and 90.2% of cases, respectively, compared to 93.2% for the EM physician. Moderate agreement between the artificial intelligence (AI) tools and the EM physician was observed (kappa: 0.476 for ChatGPT and 0.414 for Copilot).

Conclusion: LLM-based generative AI tools show promise as clinical decision support systems, enhancing diagnostic accuracy and assisting less-experienced clinicians. However, they should complement, not replace, human expertise in emergency settings.

Keywords: ChatGPT, Copilot, LLMs, generative artificial intelligence, differential diagnosis, emergency

Introduction

The diagnostic process relies on four essential components: medical history, physical examination, differential diagnoses, and diagnostic tests. Experienced clinicians have historically been able to diagnose accurately in 70%, 90% of cases by taking a thorough and detailed medical history (1,2). This demands extensive medical knowledge, sharp observational skills, and rigorous logical reasoning. The same precision is required when formulating a comprehensive list of differential diagnoses, as any error in judgment can derail the entire diagnostic process, and lead to a delayed or missed diagnosis with potentially serious consequences for patient outcomes. Unfortunately, approximately 7.4 million (5.7%) emergency department (ED) visits in the USA involve at least one diagnostic error annually causing 371,000 (0.3%) serious harms (3). A total of 89% of these errors are attributed to failures in clinical decision-making or judgment, regardless of the underlying disease. Cognitive errors causing delayed or missed diagnosis that are related to human factors such as clinical expertise, inadequate knowledge, or critical thinking have been reported in several studies (4-7).

A potential solution for mitigating cognitive errors in the diagnostic process is employing artificial intelligence (AI) technologies. Brown et al. (8) reported that improved clinical

Corresponding Author: Banu Arslan MD, University of Health Sciences Türkiye, Başakşehir Çam and Sakura City, Clinic of Emergency Medicine, İstanbul, Türkiye
E-mail: dr.banuarslan@gmail.com **ORCID ID:** orcid.org/0000-0003-0982-5351

Cite this article as: Arslan B, Erdem B, Osoydan Satıcı M, Nuhoğlu Ç, Soylu HY. Large language models in generating differential diagnoses in the emergency department; a comparative study of ChatGPT, Copilot, and emergency physician. Eurasian J Emerg Med. 2025;24(3): 201-7.

Received: 11.02.2025
Accepted: 01.05.2025
Epub: 07.07.2025
Published: 10.09.2025

decision-making and reduced risk of patient harm could be achieved in emergency patients by employing AI technologies, including AI-based symptom checkers, natural language processing tools to generate differential diagnoses, and real-time electrocardiography and X-ray interpretation tools. Similarly, Harada et al. (9) and Schwitzguebel et al. (10) reported that less-experienced physicians could enhance their diagnostic accuracy by using AI technologies in history taking and generating differential diagnoses.

With the launch of Chat Generative Pre-trained Transformer (ChatGPT) (OpenAI, San Francisco, CA) in November 2022 and Copilot (formerly Bing AI) (Microsoft, Redmond, WA) in February 2023, AI technologies became easier for the general public to use and more accessible. Their ability to comprehend inputs and generate human-like, fluent text outputs to queries, quickly drew attention across various fields, including healthcare. The latest versions of these models were built on the Generative Pre-trained Transformer-4 (GPT-4) architecture, which utilizes a transformer-based neural network to predict the next token in a document (11). GPT-4's advanced capabilities make it particularly useful in healthcare, where it can assist health professionals with tasks such as medical diagnosis and generating differential diagnosis lists. A user can input a patient's clinical manifestations into these large language model (LLM)-based generative AI tools, allowing the models to analyze medical text data and suggest potential differential diagnoses. In this research, we aimed to determine the diagnostic accuracy of these novel AI tools using real-life clinical vignettes and to have them list differential diagnoses. We believe that this approach can serve as a clinical decision support system (CDSS) tool, providing valuable assistance to less experienced health professionals in their practice in the future.

Materials and Methods

Study Setting and Data Collection

The present study was designed as a single-center, retrospective and cross-sectional study and conducted with 468 real-world clinical vignettes. The data for the 468 patients were obtained from the previous research titled "evaluating LLM-based generative AI tools in a five-level emergency triage system: a comparative study of ChatGPT Plus, Copilot Pro, and triage nurses". That study was conducted in an ED of a large urban academic hospital over a one-week period between December 11 and December 18, 2023. Adult patients were enrolled during random 24-hour intervals. Exclusions included minors, trauma cases, and incomplete data. In the triage area, nurses assessed patients while the emergency physician observed them and then documented standardized clinical vignettes.

For our research, Institutional Review Board approval was obtained University of Health Sciences Türkiye, Şişli Hamidiye Etfal Training and Research Hospital Health Practice and Research Clinical Research Ethics Committee (decision number: 4640, date: 03.12.2024).

Outcome and Procedure

The primary outcome of the study was determining the accuracy of LLM-based AI applications in generating primary and differential diagnoses for non-trauma adult patients presenting to the ED. To do so, we simultaneously introduced each clinical vignette to the GPT-4 based generative AI tools and an experienced emergency medicine (EM) physician, by asking "Can you list 5 possible diagnoses, ordered from most likely to less likely based on the presented information above?" The first diagnosis was accepted as the primary diagnosis and the others as differential diagnoses. These diagnosis lists were then evaluated by an academic EM physician and benchmarked against actual discharge diagnoses. The accuracy of LLM-based generative AI tools and EM specialists in identifying the correct diagnosis was investigated by determining the position of the final diagnosis in each list.

LLMs Based-generative AI tools

Two LLM-based generative AI tools were used in this study; ChatGPT Plus and Copilot Pro. ChatGPT Plus was accessed via OpenAI's GPT-4 interface (version dated: 11.12.2023), and Copilot Pro was used via Microsoft Copilot with GPT-4 integration as of the same date. These tools can be considered advanced AI systems that can create high-quality content, such as text, audio, code, images, and videos, based on the data they were trained on. They can analyze raw data and identify underlying patterns and structures, allowing them to generate the most statistically likely outputs in response to specific prompts. Both models ran on Open AI's GPT-4 and performed various tasks. However, its infrastructure is not publicly disclosed.

Statistical Analysis

Statistical analyses were performed using SPSS software, version 28.0. Descriptive statistics, including means, standard deviations, medians, minimum and maximum values, frequencies, and percentages, were used to summarize and characterize the data. The distribution of variables was checked with the Kolmogorov-Smirnov test. The Friedman test and the Wilcoxon test were used for the repeated measurement analysis. The chi-square test was used for comparison of qualitative data. Cohen's kappa was used to assess the level of agreement between LLMs and EM physicians.

Results

A total of 468 clinical vignettes were included in the study. Patients ranged in age from 18 to 100 years, with a mean age of 46±19.8 years and a median age of 44.5 years. The gender distribution was balanced, with 54.7% female (n=256) and 45.3% male (n=212). The majority presented with ear, nose, and throat (ENT) symptoms (15.4%, n=72), followed by respiratory (15.2%, n=71) and cardiovascular complaints (13.9%, n=65). In terms of comorbidities, 63.7% (n=298) had no comorbidities, while 36.3% (n=170) had at least one. The most common comorbidities were hypertension (16.2%, n=76), diabetes mellitus (9.6%, n=45), and coronary artery disease (6.6%, n=31). Most patients (91.2%, n=427) were discharged; 4.5% required gastrointestinal (GI) hospitalization, 3.2% were admitted to the intensive care unit (ICU), 0.9% refused care, and 0.2% died. Demographic data were presented in Table 1.

The accuracy of the EM physician, ChatGPT, and Copilot in predicting the correct diagnosis within the top five listed diagnoses was 98.5%, 97.4%, and 94.9%, and the top three listed diagnoses was 93.2%, 91.9%, and 90.2%, respectively. The EM physician identified the definitive diagnosis as the first choice in 379 cases, with an accuracy rate of 81.0%, compared to ChatGPT's 351 cases (75%) and Copilot's 345 cases (73.7%) (Table 2).

The EM physician failed to include the final diagnosis within the top five differential diagnoses in 7 cases, compared to 12 missed cases by ChatGPT and 24 by Copilot. Among clinical systems, the EM physician missed 1 ENT case, 2 GI cases, 1 musculoskeletal case, 1 genitourinary (GU) case, and 2 hematologic cases. ChatGPT missed 2 ENT, 1 cardiovascular, 1 GI, 3 musculoskeletal, 1 GU, 3 hematologic, and 1 psychiatric case. Copilot missed 4 ENT, 2 cardiovascular, 4 GI, 9 musculoskeletal, 2 hematologic, and 3 psychiatric cases. In terms of patient outcomes, the EM physician missed 5 discharged cases and 2 ward admissions, with no missed ICU or surgical/intervention cases. ChatGPT missed 9 discharged cases, 2 ward admissions, and 1 ICU admission. Copilot missed 17 discharged cases, 4 ward admissions, 3 cases requiring procedural or surgical intervention, but no ICU admissions (Table 3).

ChatGPT and the EM physician agreed on the rank of the 375 cases; 337 were identified as the first choice. Similarly, Copilot and the EM physician agreed on 361 cases, with 329 being identified as the first choice. There was a moderate agreement between all raters with Cohen's kappa values for the EM physician versus ChatGPT and for the EM physician versus Copilot being 0.476 and 0.414, respectively (p=0.000) (Table 4).

Table 1. Patient demographics, chief complaints, chronic medical conditions, and patient outcomes

		Minimum-Maximum			Median	Mean ± SD/n-%		
Age		18.0	-	100.0	44.5	46.2	±	19.8
Gender	Female					256		54.7%
	Male					212		45.3%
Chief complaints	ENT-mouth, throat, neck					72		15.4%
	Respiratory					71		15.2%
	Cardiovascular					65		13.9%
	Neurological					61		13.0%
	GI					58		12.4%
	Musculoskeletal (limp/joint pain, neck pain)					35		7.5%
	Ophthalmology					23		4.9%
	GU					20		4.3%
	Mental health					18		3.8%
	ENT- nose/ear					18		3.8%
	Skin					17		3.6%
	Fever					8		1.7%
	Poisoning					2		0.4%

Table 1. Patient demographics, chief complaints, chronic medical conditions, and patient outcomes				
		Mean ± SD/n-%		
Comorbidities	None	298		63.7%
	At least 1	170		36.3%
	At least 2	77		16.5%
	3 or more	41		8.8%
	HT	76		16.2%
	DM	45		9.6%
	CAD	31		6.6%
	Malignancy	22		4.7%
	COPD	14		3.0%
	CHF	13		2.8%
	CVD	13		2.8%
	CRF	12		2.6%
	Asthma	10		2.1%
	Migraine	9		1.9%
	Epilepsy	7		1.5%
	Others	30		6.4%
Outcome	Discharge	427		91.2%
	Hospitalization	21		4.5%
	ICU	15		3.2%
	Mortality	1		0.2%
	Refusal of care	4		0.9%

ENT: Ear, nose, and throat, SOB: Shortness of breath, HT: Hypertension, GI: Gastrointestinal, GU: Genitourinary, DM: Diabetes mellitus, CAD: Coronary artery disease, COPD: Chronic obstructive pulmonary disease, CHF: Congestive heart failure, CVD: Cerebrovascular disease, CRF: Chronic renal failure, ICU: Intensive care unit, SD: Standard deviation

Table 2. Comparison of diagnostic rankings among emergency medicine physician, ChatGPT, and Copilot							
		EM physician		ChatGPT		Copilot	
		n	%	n	%	n	%
Diagnostic rank	1	379	81.0	351	75	345	73.7
	2	37	7.9	51	10.9	49	10.5
	3	20	4.3	28	6.0	28	6.0
	4	15	3.2	14	3.0	11	2.4
	5	10	2.1	12	2.6	11	2.4
	≥6	7	1.5	12	2.6	24	5.1

EM: Emergency medicine, ChatGPT: Chat Generative Pre-trained Transforme. This table compares the performance of an emergency medicine physician, ChatGPT, and Copilot in ranking the correct diagnosis among a list of potential diagnoses, with the rankings shown from 1 (most accurate) to 5 (least accurate)

Table 3. Number of missed diagnoses across clinical systems and patient outcomes, as identified by the EM physician, ChatGPT, and Copilot. A missed case was defined as the failure to include the final diagnosis within the top five differential diagnoses generated for each patient vignette

Systems	EM physician	ChatGPT	Copilot
ENT	1	2	4
Cardiovascular system	-	1	2
GI system	2	1	4
Musculoskeletal system	1	3	9
GU system	1	1	
Hematology	2	3	2
Mental health	-	1	3
Patient outcomes			
Discharged	5	9	17
Procedural/surgical intervention	-	-	3
Admitted to a ward	2	2	4
ICU	-	1	-

ENT: Ear, nose, and throat, EM: Emergency medicine, GI: Gastrointestinal, GU: Genitourinary, ICU: Intensive care unit, ChatGPT: Chat Generative Pre-trained Transforme

Table 4. Kappa accuracy test between EM physician and LLMs

		Emergency medicine physician						Accuracy	p value
Diagnostic rank		**1**	**2**	**3**	**4**	**5**	**≥6**		
ChatGPT	1	**337**	8	2	4	0	0	80.1%	Kappa: 0.476 p=0.000
	2	22	**20**	3	3	2	1		
	3	9	6	**7**	3	3	0		
	4	5	3	2	**3**	1	0		
	5	5	0	1	2	**3**	1		
	≥6	1	0	5	0	1	**5**		
Copilot	1	**329**	12	2	2	0	0	77.1%	Kappa: 0.414 p=0.000
	2	21	**17**	6	3	2	0		
	3	14	5	**5**	3	1	0		
	4	5	1	2	**2**	1	0		
	5	5	1	0	2	**2**	1		
	≥6	5	1	5	3	4	**6**		

EM: Emergency medicine, LLMs: Large language models, ChatGPT: Chat Generative Pre-trained Transforme

Discussion

This study provides an in-depth analysis of the diagnostic performance of the GPT-4-based ChatGPT and Copilot using real-world patient data. Our findings indicate that both tools exhibit a high level of accuracy in predicting the correct diagnosis based on patients' clinical history and vital parameters. This supports the growing body of evidence suggesting that AI-driven models can enhance diagnostic accuracy in clinical settings.

Previous studies have also reported promising results regarding AI's ability to assist in diagnosis. For example, Levine et al. (12) evaluated a GPT-3-based AI model (an earlier version of ChatGPT) using 48 clinical vignettes and found that the model identified the correct diagnosis in the top three differential diagnoses with an accuracy of 88%. Similarly, another study using 30 clinical vignettes reported an accuracy rate of 83.3% in generating the correct diagnosis within a list of five possible differentials (13). More recent research conducted by Hirosawa et al. (14) has demonstrated the superiority of newer AI models, including ChatGPT-3.5 and ChatGPT-4.0. In that study, ChatGPT-3.5 achieved a 65% accuracy rate in listing the correct diagnosis within the top five differential diagnoses, while ChatGPT-4.0 improved this rate to 81%. In line with these previous studies, our research highlights the notable diagnostic performance of GPT-based

LLMs in an emergency setting. Both tools demonstrated a high degree of accuracy when listing differential diagnoses, achieving rates above 90% for the top three predictions. This improvement in performance over earlier versions of these models can be attributed to advances in model architecture, increased training data, and enhanced fine-tuning.

Notably, some of the missed diagnoses we identified involved potentially serious conditions, such as cardiovascular or hematologic disorders, which could delay critical interventions. Specifically, Copilot failed to include the correct diagnosis in three cases, that ultimately required procedural or surgical treatment, two of which were acute limb ischemia. Similarly, ChatGPT missed a diagnosis in a patient later admitted to the ICU, indicating that diagnostic oversights by AI tools could have meaningful clinical consequences. These findings highlight the importance of cautious integration of generative AI in emergency decision-making and underscore the continued need for clinical oversight.

Another notable finding in our study is the moderate agreement between the LLM-based tools and the EM physician, as reflected by the Cohen's kappa values. While accuracy was a key measure, we also evaluated this parameter to explore the potential of AI tools as supportive aids in clinical decision-making. Continuous feedback is essential for improving diagnostic accuracy (15) and enabling the smooth integration of AI into clinical workflows. Although research in this area is limited, a recent study involving 392 case descriptions reported similar results, with moderate agreement between GPT-4 and physicians (Cohen's kappa= 0.47) (0.39-0.56) in generating differential diagnoses (16). This level of agreement between EM physicians and LLM-based AI highlights their potential value, especially for supporting less-experienced clinicians in decision-making.

In light of these considerations, we propose that LLM-based generative AI tools such as ChatGPT and Copilot should be integrated into healthcare as CDSS, particularly in high-volume settings like the ED, where time constraints, high levels of stress and diagnostic complexity are prevalent. AI tools can assist in providing rapid, evidence-based suggestions, ensuring that fewer diagnoses are overlooked. This approach can help mitigate the cognitive biases that often contribute to diagnostic errors in emergency care.

Study Limitations

This study has three main limitations. First, data were collected over one week in December, a period with high rates of upper respiratory tract infections. This short duration may not fully capture seasonal variations in disease presentations and patient demographics, potentially affecting diagnostic accuracy and the study's generalizability. Second, this study used the discharge diagnosis as the final diagnosis. While discharge diagnosis serves as a reasonable benchmark, it may not always reflect the most accurate final diagnosis due to misdiagnosis. Third, our study was conducted in a controlled environment rather than real emergency settings. In real-time clinical practice, cognitive ability is influenced by multiple dynamic factors such as time constraints, high patient volume, physician workload, stress, and cognitive fatigue. These factors may affect clinical decision-making, ranking, and variety of potential diagnoses. Given these considerations, we recommend that future studies be conducted in real-time and real-world emergency settings.

Conclusion

Our study demonstrates the significant potential of LLM-based generative AI tools like ChatGPT and Copilot to assist clinicians in diagnostic reasoning. These tools should be viewed as complementary aids rather than replacements for human expertise. To ensure safe and effective integration into clinical practice, their implementation must be accompanied by continuous evaluation. While our findings are promising, future multicenter studies are needed to enhance generalizability and validate performance across diverse clinical settings.

Ethics

Ethics Committee Approval: Institutional Review Board approval was obtained University of Health Sciences Türkiye, Şişli Hamidiye Etfal Training and Research Hospital Health Practice and Research Clinical Research Ethics Committee (decision number: 4640, date: 03.12.2024).

Informed Consent: The present study was designed as a single-center, retrospective and cross-sectional study and conducted with 468 real-world clinical vignettes.

Footnotes

Authorship Contributions

Concept: B.E., Design: B.A., M.O.S., Ç.N., H.Y.S., Data Collection or Processing: B.A., B.E., Analysis or Interpretation: B.A., B.E., M.O.S., Ç.N., H.Y.S., Literature Search: B.A., B.E., Writing: H.Y.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. Peterson MC, Holbrook JH, Von Hales D, Smith NL, Staker LV. Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. West J Med. 1992;156:163-5.
2. Gruppen LD, Woolliscroft JO, Wolf FM. The contribution of different components of the clinical encounter in generating and eliminating diagnostic hypotheses. Res Med Educ. 1988;27:242-7.
3. Newman-Toker DE, Peterson SM, Badihian S, Hassoon A, Nassery N, Parizadeh D, et al. Diagnostic errors in the emergency department: a systematic review [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2022.
4. Hussain F, Cooper A, Carson-Stevens A, Donaldson L, Hibbert P, Hughes T, et al. Diagnostic error in the emergency department: learning from national patient safety incident report analysis. BMC Emerg Med. 2019;19:77.
5. Kachalia A, Gandhi TK, Puopolo AL, Yoon C, Thomas EJ, Griffey R, et al. Missed and delayed diagnoses in the emergency department: a study of closed malpractice claims from 4 liability insurers. Ann Emerg Med. 2007;49:196-205.
6. Norman G. Building on experience--the development of clinical reasoning. N Engl J Med. 2006;355:2251-2.
7. Committee on diagnostic error in health care; board on health care services; institute of medicine; the national academies of sciences, engineering, and medicine. Improving diagnosis in health care. Balogh EP, Miller BT, et al. Washington (DC): National Academies Press (US); 2015.
8. Brown C, Nazeer R, Gibbs A, Le Page P, Mitchell AR. Breaking bias: the role of artificial intelligence in improving clinical decision-making. Cureus. 2023;15:e36415.
9. Harada Y, Katsukura S, Kawamura R, Shimizu T. Efficacy of artificial-intelligence-driven differential-diagnosis list on the diagnostic accuracy of physicians: an open-label randomized controlled study. Int J Environ Res Public Health. 2021;18:2086.
10. Schwitzguebel AJ, Jeckelmann C, Gavinio R, Levallois C, Benaïm C, Spechbach H. Differential diagnosis assessment in ambulatory care with an automated medical history-taking device: pilot randomized controlled trial. JMIR Med Inform. 2019;7:e14044.
11. OpenAI. GPT-4 Technical Report. Preprint at arXiv https://doi.org/10.48550/arXiv.2303.08774 (2023).
12. Levine DM, Tuwani R, Kompa B, Varma A, Finlayson SG, Mehrotra A, et al. The diagnostic and triage accuracy of the GPT-3 artificial intelligence model: an observational study. Lancet Digit Health. 2024;6:e555-61.
13. Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic accuracy of differential-diagnosis lists generated by generative pretrained transformer 3 chatbot for clinical vignettes with common chief complaints: a pilot study. Int J Environ Res Public Health. 2023;20:3378.
14. Hirosawa T, Kawamura R, Harada Y, Mizuta K, Tokumasu K, Kaji Y, et al. ChatGPT-generated differential diagnosis lists for complex case-derived clinical vignettes: diagnostic accuracy evaluation. JMIR Med Inform. 2023;11:e48808.
15. Singh H, Connor DM, Dhaliwal G. Five strategies for clinicians to advance diagnostic excellence. BMJ. 2022;376:e068044.
16. Hirosawa T, Harada Y, Mizuta K, Sakamoto T, Tokumasu K, Shimizu T. Evaluating ChatGPT-4's accuracy in identifying final diagnoses within differential diagnoses compared with those of physicians: experimental study for diagnostic cases. JMIR Form Res. 2024;8:e59267.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.97455
Eurasian J Emerg Med. 2025;24(3): 208-13

Performance of Modified Pneumonia Severity Indexes in Community-Acquired Pneumonia

Yasemin Yılmaz[1], Aynur Yurtseven[2]

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

Abstract

Aim: In patients with community-acquired pneumonia, we observed that the impact of the age variable commonly used in Pneumonia Severity scores to predict mortality risk is disproportionate within the pneumonia severity index (PSI). In this context, our study aimed to evaluate and compare the effectiveness of pneumonia severity scores, adjusted for demographic factors such as age and gender, in predicting intensive care unit (ICU) admission and mortality, compared to PSI and CURB-65.

Materials and Methods: In 2019, data on patients diagnosed with pneumonia in the emergency department were obtained through a retrospective review. The CURB-65, PSI, age-modified PSI (mPSI), and gender- modified PSI (gmPSI) scores were calculated. The predictive performance of these scores for ICU admission and mortality was statistically analyzed.

Results: A total of 363 patients were included in the study, of whom 205 (56.4%) were male. Additionally, comparisons of the newly developed age-mPSI and gmPSI with the standard PSI and CURB-65 demonstrated no statistically significant difference in predicting 30-day mortality or ICU admission rates ($p>0.05$).

Conclusion: The findings of this study indicate that the newly developed mPSI and gmPSI scoring systems yield results comparable to those of the standard PSI scoring system.

Keywords: Community acquired pneumonia, pneumonia severity index, CURB-65, age, gender

Introduction

Community-acquired pneumonia (CAP) is a common disease with a lifetime prevalence of 20-30% in developing countries and 3-4% in developed countries (1). The mortality rate among hospitalized and diagnosed patients has been reported to vary between 4% and 21% in different settings (2). It has been reported that this rate is higher, even up to 50%, among patients admitted to the intensive care unit (ICU) (3). Treatment in patients with CAP is determined according to the general condition and prognosis of the patient, and the follow-up and treatment plan of the patient varies according to the severity of the disease (4). Therefore, various scoring systems are used to determine the prognosis. The most commonly used assessment tools are CURB-65 (confusion, blood urea nitrogen, respiratory rate, blood pressure) and Pneumonia Severity index (PSI). CURB-65 is a classification system used to predict the risk of mortality and is frequently preferred in emergency departments because it is easy to remember and contains few criteria (3). PSI is a scoring system developed by Fine et al. (5) that includes 20 parameters and is divided into 5 categories according to 30-day prognosis predictions. CURB-65 and the PSI are widely used to predict mortality in CAP (6). However, there are many studies reporting that PSI performs better than CURB-65 in predicting mortality (7,8).

Nevertheless, clinical applications have revealed certain limitations of this scoring system within the context of emergency departments (8,9). For instance, one study demonstrated that mortality and ICU admission rates were lower among patients categorized as PSI Class I/II but were paradoxically higher among those with a CURB-65 score of 1 (10). Furthermore, evidence

Corresponding Author: Yasemin Yılmaz MD, University of Health Sciences Türkiye, Ankara Atatürk Sanatorium Training and Research Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
E-mail: ysmnyaydin@gmail.com **ORCID ID:** orcid.org/0000-0003-1048-4478

Cite this article as: Yılmaz Y, Yurtseven A. Performance of modified pneumonia severity indexes in community-acquired pneumonia. Eurasian J Emerg Med. 2025;24(3): 208-13.

Received: 11.02.2025
Accepted: 09.05.2025
Epub: 07.07.2025
Published: 10.09.2025

suggests that approximately 27% of patients in PSI Class I-III may necessitate ICU admission and 40% of patients classified as low-risk were hospitalized (11,12). Although elevated scores in such systems are typically indicative of more severe illness, it is important to note that younger to middle-aged patients with low scores may also require ICU-level care. In particular, the age-related risk factor incorporated into the PSI scoring system may result in disproportionately high scores among older patients while potentially underestimating severity in younger individuals.

Many studies in the literature have shown that sex hormones play an important role in the immune response and that women have a stronger immune response than men. In general, women have a stronger response to pathogens and produce higher levels of interferon and antibodies (13,14). We aimed to assess the influence of age and sex on the performance of the PSI scoring system in this study. In particular, the impact of the newly developed age- and sex-modified, modified-PSI (mPSI) and gender- modified PSI (gmPSI) on clinical outcomes, including mortality and ICU admission rates, was analyzed.

Materials and Methods

This retrospective study was conducted in the emergency department of a tertiary hospital located in a provincial center that experiences approximately 420,000 patient visits annually. It received ethical approval from the University of Health Sciences Türkiye, Dışkapı Yıldırım Beyazıt Training and Research Hospital Clinical Research Ethics Committee (decision number: 103/13, date: 25.01.2021). Our research was designed in accordance with the Standards for Reporting of Diagnostic Accuracy Studies statement (15).

Information was extracted from electronic medical records and patient charts. A retrospective chart review was conducted by two emergency medicine specialists, each possessing a minimum of three years of experience.

Between January 1, 2019, and December 31, 2019, patients were diagnosed with pneumonia in the emergency medicine clinic. All patients over the age of 18 years with a diagnosis of pneumonia were included in the study. The inclusion criteria were based on ICD-10 codes (International Statistical Classification of Diseases, J15 and its subcodes, J15.8, J15.9) retrieved from the hospital's data system. Individuals with a history of hospitalization within the preceding seven days, a diagnosis of cystic fibrosis, tuberculosis, or Human immunodeficiency virus, a documented history of concomitant pulmonary embolism, or a history of trauma at the time of admission, as well as those with incomplete data in their records, were excluded from the study. Additionally, patients with high mortality at the time of admission and those diagnosed with pneumonia were not included in the study.

Patients diagnosed with pneumonia were initially stratified according to the CURB-65 score, the PSI, and the criteria for ICU admission (1). The PSI scoring system was subsequently modified by excluding the age and gender variables. To adjust for age-related risk, 10 points were allocated to patients aged 65 years and older, whereas no points were assigned to those under 65 years. This revised scoring model was termed the mPSI.

In the calculation of the gmPSI, female patients aged 65 years and older were assigned 10 points, while male patients in the same age group were allocated 20 points. No points were assigned to either male or female patients under the age of 65.

The predictive performance of the scoring systems (PSI, mPSI, and gmPSI) in estimating hospitalization and mortality outcomes was systematically evaluated. Patients were subsequently categorized into two distinct age groups: younger adults (18-64 years) and older adults (≥65 years), enabling a comprehensive comparison between these cohorts.

Statistical Analysis

Statistical analyses were conducted using the SPSS 22.0 software package. The normality of continuous variables was assessed through the Kolmogorov-Smirnov test, supplemented by descriptive statistical measures of Skewness and Kurtosis.

Descriptive statistics were presented as frequencies and percentages for categorical variables, while continuous variables were summarized using medians and (interquartile range 25^{th}-75^{th} percentiles). For the analysis of differences in numerical variables between two independent groups, the Mann-Whitney U test was employed. Comparisons of proportions between independent groups were conducted using chi-square analysis.

ROC analyses were conducted to evaluate the predictive efficacy of continuous variables in estimating 30-day mortality and the need for intensive care. A p value of <0.05 was considered the threshold for statistical significance.

Results

A total of 450 patients were initially enrolled in the study; however, 87 were excluded due to incomplete data. Among the remaining participants, 205 (56.4%) were male. Hospitalization rates did not differ significantly between male and female patients ($p=0.248$).

The stratified pneumonia scores based on 30-day mortality and ICU requirements for all patients are presented in Table 1. Patients who died or required ICU admission were older and

exhibited higher scores across all measurements ($p<0.05$ for all comparisons) (Table 1).

ROC analysis was conducted to identify the threshold values of pneumonia scores distinguishing between deceased and surviving patient groups. Among the scores, PSI demonstrated the highest area under the curve (AUC), while CURB-65 exhibited the lowest. All AUC values were statistically significant ($p<0.001$ for all comparisons) (Table 2, Figure 1A).

In the ROC analysis conducted to determine the threshold values of pneumonia scores differentiating patients admitted to the ICU from those not admitted, the mPSI demonstrated the highest AUC, while the CURB-65 showed the lowest. All AUC values were statistically significant ($p<0.001$ for all comparisons) (Table 3, Figure 1B).

The best cut-off values, along with the corresponding sensitivity and specificity of the PSI, mPSI, and gmPSI scores for predicting mortality and ICU admission, are presented in Table 4. The sensitivity and specificity values for all three scoring systems were found to be comparable (Table 4).

Patients were categorized into two age groups: 18-64 years (young) and ≥65 years (geriatric). In the ROC analysis conducted to identify the threshold values of pneumonia scores for predicting mortality within the young age group, the PSI demonstrated the highest AUC, whereas CURB-65 exhibited the lowest. All AUC values were statistically significant ($p<0.001$) (Table 5, Figure 2A).

In the ROC analysis conducted to determine the threshold values of pneumonia scores for predicting mortality in the geriatric group, PSI and mPSI exhibited the highest AUC values, while CURB-65 demonstrated the lowest (Table 6, Figure 2B).

In the ROC analysis conducted to identify the threshold values of pneumonia scores for predicting ICU admission in the young patient group, the highest AUC value was observed for the mPSI, while the lowest was associated with the PSI (Table 7, Figure 3A).

In the ROC analysis conducted to determine the threshold values of pneumonia scores for intensive care admission in the geriatric population, the highest AUC value was observed for the PSI score, while the lowest was associated with the CURB-65 score (Table 8, Figure 3B).

Discussion

Treatment, prognosis, and care location decisions are often facilitated by scoring systems, which, despite inherent challenges and limitations, serve as valuable tools in the clinical decision-making process. These systems primarily aim to minimize unnecessary hospital admissions while ensuring appropriate patient placement and management. The PSI, a widely utilized scoring system, comprises 20 variables encompassing patient age, history of nursing home residence, comorbid conditions, vital signs, laboratory results, chest radiographic findings, and oxygenation status. The assessment is conducted in two distinct stages. In this scoring system, the age variable is directly assigned based on the numerical age in male patients, while, for female patients, it is calculated as age minus 10. Consequently, this approach may result in low scores for young patients with severe respiratory failure in the absence of comorbidities, while generating high scores for very elderly patients. Such discrepancies can create confusion for clinicians and potentially lead to inappropriate hospitalization and treatment decisions.

Table 2. AUC values of pneumonia scores according to 30-day mortality

	AUC	Confidence interval	p value
CURB-65	0.700	0.637-0.763	<0.001
PSI	0.803	0.750-0.855	<0.001
mPSI	0.794	0.740-0.849	<0.001
gmPSI	0.793	0.739-0.847	<0.001

AUC: Area under the curve, PSI: Pneumonia severity index, mPSI: Modified-PSI, gmPSI: Gender-mPSI

Table 1. Pneumonia severity scores stratified by 30-day mortality and intensive care unit admission requirement

	Mortality in 30 days			ICU hospitalization		
	Exitus	Alive	p value	None ICU admission	ICU admission	p value
Gender n (%) Female	33 (44.6%)	125 (43.3%)	0.835	106 (41.4%)	52 (48.6%)	0.248
Age (years) (min-max)	80 (70-86)	75 (63-83.5)	0.005	75 (63-81.7)	80 (70-87)	<0.001
CURB-65 median (min-max)	2 (2-3)	2 (1-2)	<0.001	2 (1-2)	2 (2-3)	<0.001
PSI score	164 (138-191)	119 (93-149)	<0.001	112.5 (91-140)	163(142-180)	<0.001
mPSI, median (min-max)	100 (80-120)	60 (40-80)	<0.001	50 (30-80)	100 (80-120)	<0.001
gmPSI, median (min-max)	100 (80-120)	60 (40-90)	<0.001	60 (40-80)	100 (90-120)	<0.001
ICU admission n (%)	55 (74.3%)	58 (20.1%)	<0.001	-	-	-

ICU: Intensive care unit, PSI: Pneumonia severity index, mPSI: Modified-PSI, gmPSI: Gender-mPSI

The literature contains numerous studies evaluating the sensitivity and specificity of CURB-65 and PSI scores in predicting mortality and the necessity for ICU admission (16-18). In a study conducted in Iran, CURB-65 demonstrated higher accuracy in predicting mortality and the need for intensive care hospitalization in patients with CAP (17). In contrast, studies conducted by Akça et al. (17) and Satici et al. (18) found that PSI exhibited superior performance in predicting mortality.

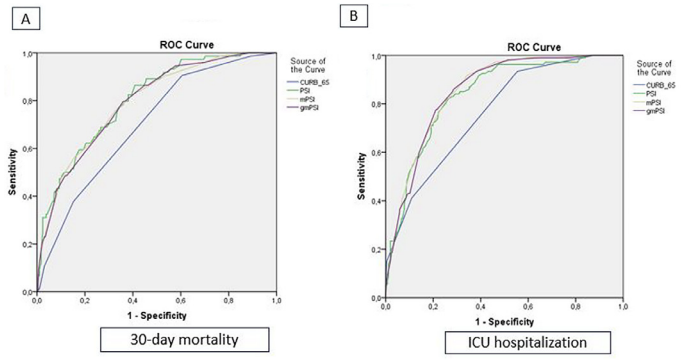


Figure 1. ROC analysis of pneumonia scores according to 30-day mortality or ICU hospitalization for all patients. A) ROC analysis of pneumonia scores according to 30-day mortality for all patients, B) ROC analysis of pneumonia scores according to ICU hospitalization for all patients

ICU: Intensive care unit

Table 3. AUC values of pneumonia scores by ICU

	AUC	Confidence interval	p value
CURB-65	0.764	0.714-0.815	<0.001
PSI	0.839	0.796-0.882	<0.001
mPSI	0.856	0.818-0.895	<0.001
gmPSI	0.854	0.815-0.892	<0.001

AUC: Area under the curve, ICU: Intensive care unit, PSI: Pneumonia severity index, mPSI: Modified-PSI, gmPSI: Gender-mPSI

Table 4. Threshold values and sensitivity and specificity of pneumonia scores

	PSI	mPSI	gmPSI
Mortality			
Best cut-off point	121%	57%	52%
Sensitivity	89%	90%	95%
Specificity	55%	47%	42%
ICU			
Best cut-off point	123%	62%	62%
Sensitivity	90%	92%	94%
Specificity	62%	66%	62%

ICU: Intensive care unit, PSI: Pneumonia severity index, mPSI: Modified-PSI, gmPSI: Gender-mPSI

In a retrospective study involving 1,902 patients, Zhang et al. (19) categorized the patients into three age groups: 18-64 years, 65-84 years, and ≥85 years. Although PSI Class III is considered a low-risk category for mortality, they observed a mortality rate of 7.5% in the 18-64 age group, compared to 2.1% in the 65-84 age group and 2.6% in those aged ≥85 years. Accordingly, they suggested that PSI Class III may not be appropriate for classification as a low mortality risk in younger patients. They reported that while

Table 5. AUC values of pneumonia scores according to 30-day mortality at 18-64 years of age

	AUC	Confidence interval	p value
CURB-65	0.760	0.594-0.927	0.008
PSI	0.870	0.756-0.985	<0.001
mPSI	0.825	0.688-0.963	=0.001
gmPSI	0.825	0.687-0.962	=0.001

AUC: Area under the curve, PSI: Pneumonia severity index, mPSI: Modified-PSI, gmPSI: Gender-mPSI

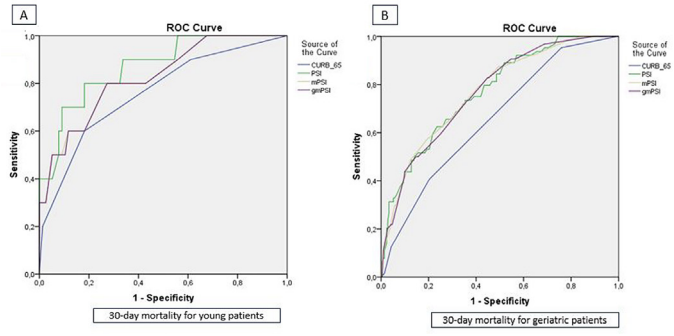


Figure 2. ROC analysis of pneumonia scores according to 30-day mortality for young and geriatric patients. A) ROC analysis of pneumonia scores according to 30-day mortality for young patients, B) ROC analysis of pneumonia scores according to 30-day mortality for geriatric patients

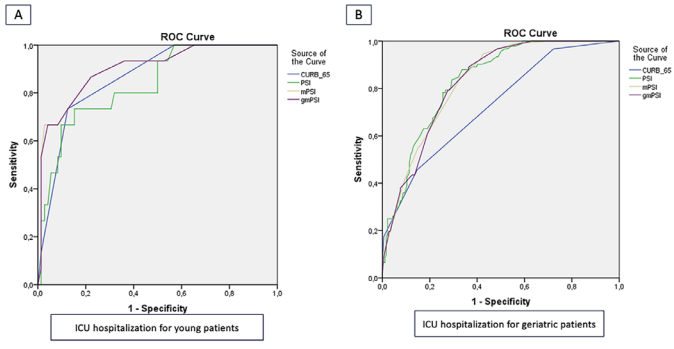


Figure 3. ROC analysis of pneumonia scores according to ICU hospitalization for young and geriatric patients. A) ROC analysis of pneumonia scores according to ICU hospitalization for young patients, B) ROC analysis of pneumonia scores according to ICU hospitalization for geriatric patients

ICU: Intensive care unit

Table 6. AUC values of pneumonia scores for predicting 30-day mortality in patients aged ≥65 years

	AUC	Confidence interval	p value
CURB-65	0.658	0.585-0.731	<0.001
PSI	0.771	0.709-0.834	<0.001
mPSI	0.771	0.708-0.834	<0.001
gmPSI	0.769	0.707-0.832	<0.001

AUC: Area under the curve, PSI: Pneumonia severity index, mPSI: Modified-PSI, gmPSI: Gender-mPSI

Table 7. AUC values of pneumonia scores for 18-64 years according to ICU

	AUC	Confidence interval	p value
CURB-65	0.865	0.775-0.955	<0.001
PSI	0.835	0.726-0.944	<0.001
mPSI	0.901	0.814-0.988	<0.001
gmPSI	0.900	0.814-0.987	<0.001

AUC: Area under the curve, ICU: Intensive care unit, PSI: Pneumonia severity index, mPSI: Modified-PSI, gmPSI: Gender-mPSI

Table 8. AUC values of pneumonia scores for predicting intensive care unit admission in the geriatric population

	AUC	Confidence interval	p value
CURB-65	0.728	0.666-0.790	<0.001
PSI	0.833	0.787-0.880	<0.001
mPSI	0.830	0.783-0.876	<0.001
gmPSI	0.828	0.781-0.875	<0.001

AUC: Area under the curve, ICU: Intensive care unit, PSI: Pneumonia severity index, mPSI: Modified-PSI, gmPSI: Gender-mPSI

PSI was more sensitive than CURB-65 in predicting mortality, it misclassified the low-risk group among younger patients, and the discriminative power of both CURB-65 and PSI diminished with increasing age (19). In a prospective study involving 987 patients, Chen et al. (20) categorized the patients into age groups as defined in the study and observed mortality rates of 20% in the 18-64 age group, 11% in the 65-84 age group, and 11.5% among those aged 85 years and over. They demonstrated that 77.8% of the patients who died in the young adult group were categorized as low-risk, a significantly higher proportion compared to those categorized as low-risk in the elderly and very old groups. They attributed the poor performance of the PSI score in elderly patients to the significant influence of the age variable and proposed that applying a non-linear adjustment to the age variable would enhance the PSI score's accuracy. Simonetti et al. (21) also observed in their prospective study that the effectiveness of PSI and CURB-65 in predicting mortality diminished with advancing age. In our clinical practice, we found it necessary to critically evaluate the age component of the PSI score. Our observations indicated that some very elderly patients did not require intensive care despite having a high PSI score, whereas certain younger patients required intensive care despite a low PSI score.

In our study, a novel scoring system was developed and implemented to address the confusion caused by existing scoring systems for patients requiring hospitalization in the ward or ICU, or those who can be managed as outpatients. Based on the data obtained from these scoring systems, we found that the highest AUC value for predicting 30-day mortality was associated with PSI, while the lowest was observed with CURB-65.

Comparisons according to the patients' age groups indicated that the highest AUC value for 30-day mortality in the 18-64 age group was related to PSI, and the highest AUC value for the need for ICU admission was related to mPSI. We found that the highest AUC value for 30-day mortality in the geriatric age group was related to both PSI and mPSI, and the highest AUC value for the need for ICU admission was related to PSI. PSI scoring was more significant than mPSI scoring in terms of the need for intensive care admission in the geriatric age group. However, the results of both scores were similar in the geriatric age group in terms of mortality. In the 18-64 age group, the significance of mPSI scoring in intensive care admissions led to the conclusion that the age factor had an effect on determining the need for admission. However, after a more detailed examination of the findings, we found that there was no significant difference between the scoring systems.

Study Limitations

Our study has a retrospective design, and the consequent data loss is one of our limitations. There may be missing data due to incorrect diagnosis code records. Due to the retrospective nature of the study, certain variables like confusion may have been under-documented in the medical records, which could have resulted in lower CURB-65 and PSI scores. There are limitations in our study because patient data were collected in a single center and the study was conducted with a specific sample size.

Conclusion

In our study, we have demonstrated that the mPSI and gmPSI scoring systems, can reduce hospital burden and additional costs by decreasing unnecessary intensive care admissions, similar to PSI. Additionally, they exhibit comparable sensitivity to PSI, which is used as the standard for predicting early mortality. Therefore, we believe that its application in both elderly and younger patients could facilitate a more objective assessment.

Ethics

Ethics Committee Approval: It received ethical approval from the University of Health Sciences Türkiye, Dışkapı Yıldırım Beyazıt Training and Research Hospital Clinical Research Ethics Committee (decision number: 103/13, date: 25.01.2021).

Informed Consent: Retrospective study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: Y.Y., A.Y., Concept: Y.Y., A.Y., Design: Y.Y., A.Y., Data Collection or Processing: Y.Y., A.Y., Analysis or Interpretation: Y.Y., A.Y., Literature Search: Y.Y., A.Y., Writing: Y.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. Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, et al. Infectious Diseases Society of America/American thoracic society consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis. 2007;44(Suppl 2):S27-72.
2. Shah BA, Ahmed W, Dhobi GN, Shah NN, Khursheed SQ, Haq I. Validity of pneumonia severity index and CURB-65 severity scoring systems in community acquired pneumonia in an Indian setting. Indian J Chest Dis Allied Sci. 2010;52:9-17.
3. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 2003;58:377-82.
4. Kara PH, Demircan A, Akinci E, Bildik F, Aygencel G, Ozsarac M. Focal mass-like opacity on chest radiography: round pneumonia. J Emerg Med. 2010;39:e89-90.
5. Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. N Engl J Med. 1997;336:243-50.
6. Oktariani Nur R, Pitoyo Ceva W, Singh Gurmeetb, Mansjoer Arifa. CURB 65 score as a predictor of early mortality in hospital-acquired pneumonia. The Egyptian Journal of Chest Diseases and Tuberculosis. 2019;68:231-5.
7. Aujesky D, Auble TE, Yealy DM, Stone RA, Obrosky DS, Meehan TP, et al. Prospective comparison of three validated prediction rules for prognosis in community-acquired pneumonia. Am J Med. 2005;118:384-92.
8. İlhan G, Karakaya Z, Akyol PY, Topal FE, Payza U. Evaulation of PSİ and CURB-65 scoring systems which patients were diagnosed as community-acquired pneumonia in emergency department. CMJ. 2017;39:586-96.
9. Olcay HÖ, Çorbacıoğlu ŞK, Emektar E, Çevik Y. Effectiveness of the YEARS algorithm added to classical clinical decision-making rules in suspected pulmonary embolism. Glob Emerg Crit Care. 2024;3:58-62.
10. Ananda-Rajah MR, Charles PG, Melvani S, Burrell LL, Johnson PD, Grayson ML. Comparing the pneumonia severity index with CURB-65 in patients admitted with community acquired pneumonia. Scand J Infect Dis. 2008;40:293-300.
11. Rosón B, Carratalà J, Dorca J, Casanova A, Manresa F, Gudiol F. Etiology, reasons for hospitalization, risk classes, and outcomes of community-acquired pneumonia in patients hospitalized on the basis of conventional admission criteria. Clin Infect Dis. 2001;33:158-65.
12. Angus DC, Marrie TJ, Obrosky DS, Clermont G, Dremsizov TT, Coley C, et al. Severe community-acquired pneumonia: use of intensive care services and evaluation of American and British Thoracic Society Diagnostic criteria. Am J Respir Crit Care Med. 2002;166:717-23.
13. Taneja V. Sex hormones determine immune response. Front Immunol. 2018;9:1931.
14. Straub RH. The complex role of estrogens in inflammation. Endocr Rev. 2007;28:521-74.
15. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open. 2016;6:e012799.
16. Alavi-Moghaddam M, Bakhshi H, Rezaei B, Khashayar P. Pneumonia severity index compared to CURB-65 in predicting the outcome of community acquired pneumonia among patients referred to an Iranian emergency department: a prospective survey. Braz J Infect Dis. 2013;17:179-83.
17. Akça HŞ, Algın A, Özdemir S, Sevimli H, Kokulu K, Eroğlu SE. Comparison of the efficacy of PSI, CURB-65, CALL and BCRSS in predicting prognosis and mortality in COVID-19 patients. J Exp Clin Med. 2021;38:434-9.
18. Satici C, Demirkol MA, Sargin Altunok E, Gursoy B, Alkan M, Kamat S, et al. Performance of pneumonia severity index and CURB-65 in predicting 30-day mortality in patients with COVID-19. Int J Infect Dis. 2020;98:84-9.
19. Zhang ZX, Yong Y, Tan WC, Shen L, Ng HS, Fong KY. Prognostic factors for mortality due to pneumonia among adults from different age groups in Singapore and mortality predictions based on PSI and CURB-65. Singapore Med J. 2018;59:190-8.
20. Chen JH, Chang SS, Liu JJ, Chan RC, Wu JY, Wang WC, et al. Comparison of clinical characteristics and performance of pneumonia severity score and CURB-65 among younger adults, elderly and very old subjects. Thorax. 2010;65:971-7.
21. Simonetti AF, Viasus D, Garcia-Vidal C, Carratalà J. Management of community-acquired pneumonia in older adults. Ther Adv Infect Dis. 2014;2:3-16.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.59254
Eurasian J Emerg Med. 2025;24(3): 214-8

Reasons, Outcomes, and Risk Factors for Emergency Department Visits in Pediatric Tracheostomy Patients: A Five-Year Single-Center Study

Hanife Tuğçe Çağlar[1], Sevgi Pekcan[2], Gökçen Ünal[3], Fatih Ercan[2], Fatma Nur Ayman[2], Suat Savaş[2], Bahar Ece Tokdemir[2], Talha Kuyucuoğlu[2], Ali Batuhan Alkan[2], Özge Metin Akcan[4], Mustafa Gençeli[4]

[1]Konya City Hospital, Clinic of Pediatric Pulmonology, Konya, Türkiye
[2]Necmettin Erbakan University Faculty of Medicine, Department of Pediatric Pulmonology, Konya, Türkiye
[3]Sincan Training and Research Hospital, Clinic of Pediatric Pulmonology, Ankara, Türkiye
[4]Necmettin Erbakan University Faculty of Medicine, Department of Pediatric Infectious Diseases, Konya, Türkiye

Abstract

Aim: To evaluate the frequency, causes, and clinical outcomes of emergency department (ED) visits among pediatric patients with tracheostomies and to identify the risk factors associated with frequent ED utilization.

Materials and Methods: This retrospective case-control study included pediatric patients (0-18 years) with tracheostomies who were followed up at a tertiary care center between January 2020 and December 2024. Patients were divided into two groups based on the number of ED visits per year (<2 vs. ≥2). Demographic, clinical, and caregiver-related data were collected. Statistical analyses were conducted to identify factors associated with frequent ED visits.

Results: A total of 157 patients were included, with a median age of 28 months at the time of tracheostomy. During the 5-year period, these patients visited the ED 978 times. The most common reasons for ED visits were tracheostomy cannula replacement (37.9%) and acute respiratory failure (20.1%). Patients who visited the ED ≥2 times per year (n=31, 19.7%) were more likely to be fed orally (p=0.006), have less caregiver training in tracheostomy and device care (p=0.014 and p=0.030, respectively), and have Pseudomonas aeruginosa colonization (p=0.020). Mortality was also significantly higher in this group (32.3% vs. 11.9%, p=0.011).

Conclusion: Frequent ED visits in children with tracheostomies are associated with modifiable factors such as feeding method, caregiver training, and chronic airway colonization. Structured education programs and preventive care strategies may reduce healthcare utilization and improve patient outcomes in this vulnerable population.

Keywords: Tracheostomy, pediatric emergency department, caregiver training, respiratory complications, pseudomonas aeruginosa, frequent emergency department visits

Introduction

Tracheostomy is an invasive, life-saving procedure performed for reasons such as upper airway obstruction, prolonged mechanical ventilation, or inadequate secretion clearance (1,2). Although tracheostomy is less common in children than in adults, this intervention requires significant care, poses a high risk of complications, and has significant social effects (3,4). Children with tracheostomies may require frequent hospital visits due to acute complications, such as respiratory tract infections, cannula obstruction, and cannula dislodgement (5,6). A significant proportion of these visits occur in emergency departments (EDs).

Data on ED visits of children with tracheostomies are critical to improving the management of these patients and preventing

Corresponding Author: Hanife Tuğçe Çağlar MD, Konya City Hospital, Clinic of Pediatric Pulmonology, Konya, Türkiye
E-mail: h.tugce.s@gmail.com **ORCID ID:** orcid.org/0000-0003-1378-9250
Cite this article as: Çağlar HT, Pekcan S, Ünal G, Ercan F, Ayman FN, Savaş S, et al. Reasons, outcomes, and risk factors for emergency department visits in pediatric tracheostomy patients: a five-year single-center study. Eurasian J Emerg Med. 2025;24(3): 214-8.

Received: 18.07.2025
Accepted: 25.07.2025
Epub: 12.08.2025
Published: 10.09.2025

possible complications. However, the current literature is limited with regard to data on the causes, frequency, and outcomes of these visits (6,7). This may make it difficult for healthcare professionals to meet the needs of this special patient group (5,8).

This study aims to retrospectively examine the reasons for ED visits, clinical conditions during these visits, hospitalization rates, and the risk factors for frequent ED visits among children with tracheostomies.

Materials and Methods

Study Design

This retrospective case-control study was carried out between January 2020 and December 2024 in the Department of Pediatric Pulmonology at the Necmettin Erbakan University Faculty of Medicine. The study was approved by the Ethics Committee of Necmettin Erbakan University with (decision number: 2025/5711, date: 25.04.2025).

We retrospectively reviewed the medical records of patients aged 0-18 years with tracheostomies who were under follow-up at our clinic between January 2020 and December 2024. Patients with incomplete data and those older than 18 years were excluded from the study.

The following patient demographics were recorded: age at tracheostomy, educational status, parents' educational status, caregiver, distance from home to hospital, routine feeding method [oral, nasogastric tube, or percutaneous endoscopic gastrostomy (PEG)], and caregiver training information were recorded. Furthermore, the clinical characteristics of patients, including Pseudomonas colonization, primary cause of tracheostomy, current status (deceased or alive with tracheostomy), number of ED visits, primary diagnosis/reasons for ED visits, ED outcomes [discharged, admission to ward, and admission to pediatric intensive care unit (PICU)], and length of stay in hospital, were recorded.

The primary endpoint of the study is to determine the frequency of ED visits in the period following tracheostomy placement. Visits to other hospitals' EDs were not considered. Frequent visits were defined as two or more ED visits per year.

Statistical Analysis

SPSS 22 (IBM Corp., released in 2011) IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.) was used to analyze the data in our study. The Shapiro-Wilk test was used to determine if the variables were normally distributed. Descriptive analyses were expressed as the mean ± standard deviation for normally distributed variables and as the median (minimum-maximum) for non-normally distributed variables. Categorical data were expressed as n (%). Patients were divided into two groups based on the number of ED visits per year (<2 and ≥2). All ED visits were considered equal, regardless of the complexity of the reason for the visit (e.g., basic tube cleaning or respiratory failure). Risk factors for frequent ED visits were identified by comparing the two groups. Categorical variables were compared using the chi-squared test. The Mann-Whitney U test was used for continuous variables in comparisons between the two groups. A p value of <0.05 was accepted as the level of significance.

Results

A total of 165 children with tracheostomies who were being followed up at our clinic were to be included in the study. However, eight patients were excluded because their data could not be accessed. Finally, 157 patients were included in the study. The median (minimum-maximum) age at tracheostomy was 28 (1-210) months. Among the patients, 87 (55.4%) were male and 70 (44.6%) were female. The primary caregiver was the mother for most patients (n=145, 92.4%). The most common reason for tracheostomy was severe neurological disease (n=107, 68.2%). More than half of the patients (n=85, 54.1%) were fed via PEG. The median (minimum-maximum) number of ED visits per year for all children was 1.13 (1-11.34). Thirty-one patients (19.7%) had two or more ED visits per year. These patients were statistically significantly more likely to be fed per oral, have caregivers with less training in tracheostomy care, and have caregivers with less device training. Additionally, patients with Pseudomonas colonization visited the ED significantly more frequently ($p<0.05$). Table 1 presents the demographic data of the patients included in the study, grouped according to the number of ED visits.

During the 5-year period, 157 patients visited the ED a total of 978 times. The most common reason for ED visits was tracheostomy tube replacement (n=371, 37.9%), followed by acute respiratory failure (n=197, 20.1%) and upper respiratory tract infections/pneumonia (n=134, 13.7%). Of these patients, 599 (61.2%) were discharged, 229 (23.4%) were admitted to a ward, and 150 (15.3%) were admitted to the PICU. The median length of hospital stay for these visits was 10 days (range 1-135 days). The clinical characteristics of the patients' ED visits are shown in Table 2.

Discussion

This retrospective study examined the reasons for ED visits among pediatric patients with tracheostomies, as well as the risk factors associated with frequent visits. Over a five-year period, a total of 157 patients with tracheostomy visited the ED 978 times, with approximately one-fifth of the patients visiting two

Table 1. Comparison of patient demographics and risk factors according to emergency department visit frequency				
	<2 visits/year	**≥2 visits/year**	**Total**	**p value**
Total number of patients, n (%)	126 (80.3)	31 (19.7)	157 (100)	-
Tracheostomy age, months, median (minimum-maximum)	29 (1-210)	19 (1-182)	28 (1-210)	0.268
Male gender, n (%)	73 (57.9)	14 (45.2)	87 (55.4)	0.200
Education status, n (%) Going to school Education at home At school age but not going to school Not at school age	 7 (5.6) 10 (7.9) 61 (48.4) 48 (38.1)	 1 (3.2) 2 (6.5) 11 (35.5) 17 (54.8)	 8 (5.1) 12 (7.6) 72 (45.9) 65 (41.4)	0.403
Mother education status, n (%) Primary school graduate High school graduate University graduate	 83 (65.9) 25 (19.8) 18 (14.3)	 15 (48.4) 11 (35.5) 5 (16.1)	 98 (62.4) 36 (22.9) 23 (14.6)	0.140
Father education status, n (%) Primary school graduate High school graduate University graduate	 64 (50.8) 36 (28.6) 26 (20.6)	 15 (48.4) 9 (29) 7 (22.6)	 79 (50.3) 45 (28.7) 33 (21)	0.963
Primary caregiver, n (%) Mother Father Other family member Nurse	 114 (90.5) 8 (6.3) 3 (2.4) 1 (0.8)	 31 (100) 0 (0) 0 (0) 0 (0)	 145 (92.4) 8 (5.1) 3 (1.9) 1 (0.6)	0.362
Distance to our hospital, n (%) <10 km 10-50 km >50 km	 42 (33.3) 40 (31.7) 44 (34.9)	 13 (41.9) 12 (38.7) 6 (19.4)	 55 (35) 52 (33.1) 50 (31.8)	0.249
Transport, n (%) Private vehicle Ambulance	 94 (74.6) 32 (25.4)	 20 (64.5) 11 (35.5)	 114 (72.6) 43 (27.4)	0.259
Feeding, n (%) Per oral Via nasogastric catheter Via percutaneous endoscopic gastrostomy	 25 (19.8) 29 (23) 72 (57.1)	 14 (45.2) 8 (25.8) 9 (29)	 39 (24.8) 37 (23.6) 81 (51.6)	**0.006**
Was the caregiver fully trained in managing tracheostomy patients? n (%) Yes	 120 (95.2)	 25 (80.6)	 145 (92.4)	**0.014**
Was the caregiver fully trained on how to use the devices? n (%) Yes	 114 (90.5)	 23 (74.2)	 137 (87.3)	**0.030**
Was the caregiver fully trained in cardiopulmonary resuscitation? n (%) Yes	 103 (81.7)	 22 (71)	 125 (79.6)	0.182
Does the device have a humidifier? n (%) Yes	 94 (74.6)	 21 (67.7)	 115 (73.2)	0.439
Is there a coughing device? n (%) Yes	 19 (15.1)	 7 (22.6)	 26 (16.6)	0.314
Is there a electric generator for devices? n (%) Yes	 48 (38.1)	 12 (38.7)	 60 (38.2)	0.950
Primary cause of tracheostomy? n (%) Preterm birth Severe neurological disease Severe lung disease	 10 (7.9) 91 (72.2) 25 (19.8)	 1 (3.2) 16 (51.6) 14 (45.2)	 11 (7) 107 (68.2) 39 (24.8)	**0.013**
Pseudomonas colonization, n (%) Yes	 60 (47.6)	 22 (71)	 82 (52.2)	**0.020**
Does he/she perform regular respiratory physiotherapy? n (%) Yes	 96 (76.2)	 21 (67.7)	 117 (74.5)	0.333
Current status, n (%) Alive with tracheostomy Deceased	 111 (81.1) 15 (11.9)	 21 (67.7) 10 (32.3)	 132 (84.1) 25 (15.9)	**0.011**

Table 2. Distribution of clinical diagnoses and ed outcomes in pediatric tracheostomy patients

Primary diagnosis and reasons for visit, n (%)	
Acute gastroenteritis	91 (9.3)
Sepsis	57 (5.8)
Upper respiratory tract infection/pneumonia	134 (13.7)
Tracheostomy cannula exchange	371 (37.9)
Gastrostomy care	40 (4.1)
Epileptic seizure	53 (5.4)
Acute respiratory failure	197 (20.1)
Urinary tract infection	9 (0.9)
Bradycardia/tachycardia	18 (1.8)
Trauma	8 (0.8)
Emergency department outcomes, n (%)	
Discharged from emergency department	599 (61.2)
Ward admission	229 (23.4)
Intensive care unit admission	150 (15.3)
Length of stay in hospital, days, median (minimum-maximum)	10 (1-134)

or more times per year. The main risk factors for frequent visits were per-oral feeding, a lack of caregiver training in managing tracheostomy patients and devices, Pseudomonas colonization, and preterm birth.

A study by Tarfa et al. (6) also reported that respiratory complications and cannula-related problems were the most common reasons for ED visits among children with tracheostomies. In this study, the most common reasons for ED visits were tracheostomy cannula replacement (37.9%) and acute respiratory failure (20.1%). Of the cases admitted to the hospital, approximately 40% required the PICU, with some stays exceeding 100 days. These findings demonstrate that children with tracheostomies require significant healthcare resources. These findings underscore the significant impact of technical and clinical challenges in home care on ED visits.

The results of this study showed that frequent ED users had significantly less tracheostomy and device training. These results suggest that caregiver training plays a key role in reducing the frequent ED visits for this patient group. The American Association for Respiratory Care's clinical practice guideline states that training caregivers at home can reduce complications and limit the use of health services (5). Inadequate caregiver training may lead to delays in recognizing early warning signs of airway obstruction, secretion accumulation, or tube malfunction, all of which can result in avoidable ED visits. Clearly, including caregivers in simulation-based training, video training, or structured home care programs, as well as supporting families through training, will reduce ED visits.

This study found that severe neurological diseases were the most common cause of tracheostomy. For these patients, a combination of muscle weakness and difficulty swallowing increases the risk of aspiration. Studies showed that children fed via PEG experience fewer complications (9). The reason our patients who are fed orally visited the ED more frequently is also due to an increased risk of aspiration. Early identification of high-risk patients and timely transition to safer feeding routes may prevent recurrent respiratory complications and hospital visits. Trained caregivers also play an important role in the early identification of patients at high risk of aspiration.

The presence of Pseudomonas in the airway often leads to persistent inflammation and mucus overproduction, which may exacerbate respiratory distress and prompt frequent ED utilization. This study showed that patients with Pseudomonas aeruginosa colonization visited the ED more frequently, which can be explained by an increased risk of respiratory tract infection. Previous studies have demonstrated that Pseudomonas aeruginosa causes chronic colonization in tracheostomized patients, leading to increased hospital admissions (6,7). A large-scale study by Russell et al. (10) found that Pseudomonas aeruginosa colonization was associated with the severity of respiratory symptoms and hospital admissions.

A total of 32.3% of patients who visited the ED more than two times per year died. This rate was significantly higher than the rate of patients who visited the ED less than twice per year. These results suggest that frequent hospital visits may indicate serious underlying clinical problems. Berry et al. (8) demonstrated that frequent hospital visits among children with tracheostomies are associated with mortality. Various studies have reported higher mortality rates in children with complex care needs, such as neurological diseases, respiratory device dependence, and enteral feeding requirements (11,12).

It is well known that a significant proportion of ED visits after tracheostomy are avoidable (13). Strategies such as structured programs that support tracheostomy and device training, regular respiratory physiotherapy, and the early detection of infections may reduce repeat visits. However, numerous variables such as socioeconomic factors, caregiver education, and access to healthcare services in the area of residence influence this process (14). Although our study did not find a significant correlation between age at tracheostomy and visit frequency, the literature suggests that children who undergo tracheostomy at an early age may be more susceptible to complications and infections (15). This is thought to be influenced by factors such as home care conditions, the number of devices used, and applications of physical therapy.

Study Limitations

This study has several limitations. First, it is retrospective and single-center, which makes it difficult to evaluate referrals from

external centers. Additionally, some care factors are based on subjective reports, introducing a risk of bias. However, the study's reliability is enhanced by its relatively large patient sample size and five-year follow-up period.

Conclusion

In conclusion, this study revealed that a significant proportion of children with tracheostomies frequently visit EDs, and these visits are associated with various clinical and educational factors. Developing comprehensive care programs, expanding caregiver education, and providing increased home care support for this patient group could effectively reduce healthcare utilization and complications.

Ethics

Ethics Committee Approval: The study was approved by the Ethics Committee of Necmettin Erbakan University with (decision number: 2025/5711, date: 25.04.2025).

Informed Consent: This is retrospective study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: F.E., B.E.T., Concept: H.T.Ç., G.Ü., F.N.A., Design: H.T.Ç., F.N.A., T.K., Ö.M.A., Data Collection or Processing: H.T.Ç., S.P., G.Ü., F.E., F.N.A., S.S., B.E.T., T.K., A.B.A., Ö.M.A., M.G., Analysis or Interpretation: H.T.Ç., G.Ü., F.N.A.,B.E.T. Literature Search: H.T.Ç., G.Ü., F.N.A., A.B.A., Writing: H.T.Ç., S.P., F.E., M.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. Carron JD, Derkay CS, Strope GL, Nosonchuk JE, Darrow DH. Pediatric tracheotomies: changing indications and outcomes. Laryngoscope. 2000;110:1099-104.
2. Mahadevan M, Barber C, Salkeld L, Douglas G, Mills N. Pediatric tracheotomy: 17 year review. Int J Pediatr Otorhinolaryngol. 2007;71:1829-35.
3. Berry JG, Graham RJ, Roberson DW, Rhein L, Graham DA, Zhou J, et al. Patient characteristics associated with in-hospital mortality in children following tracheotomy. Arch Dis Child. 2010;95:703-10.
4. Funamura JL, Durbin-Johnson B, Tollefson TT, Harrison J, Senders CW. Pediatric tracheotomy: indications and decannulation outcomes. Laryngoscope. 2014;124:1952-8.
5. Volsko TA, Parker SW, Deakins K, Walsh BK, Fedor KL, Valika T, et al. AARC clinical practice guideline: management of pediatric patients with tracheostomy in the acute care setting. Respir Care. 2021;66:144-55.
6. Tarfa RA, Morris J, Melder KL, McCoy JL, Tobey ABJ. Readmissions and mortality in pediatric tracheostomy patients: are we doing enough? Int J Pediatr Otorhinolaryngol. 2021;145:110704.
7. Goldenberg D, Ari EG, Golz A, Danino J, Netzer A, Joachims HZ. Tracheotomy complications: a retrospective study of 1130 cases. Otolaryngol Head Neck Surg. 2000;123:495-500.
8. Berry JG, Graham DA, Graham RJ, Zhou J, Putney HL, O'Brien JE, et al. Predictors of clinical outcomes and hospital resource use of children after tracheotomy. Pediatrics. 2009;124:563-72.
9. Zhang Z, Akulian J, Hong Y, Liu N, Chen Y; AME Critical Care Collaborative Group. How should this patient with repeated aspiration pneumonia be managed and treated? -a proposal of the Percutaneous ENdoscopIc Gastrostomy and Tracheostomy (PENIIGhT) procedure. J Thorac Dis. 2016;8:3720-7.
10. Russell CJ, Simon TD, Mamey MR, Newth CJL, Neely MN. Pseudomonas aeruginosa and post-tracheotomy bacterial respiratory tract infection readmissions. Pediatr Pulmonol. 2017;52:1212-8.
11. Brown SES, Hall M, Cassidy RB, Zhao X, Kheterpal S, Feudtner C. Tracheostomy, Feeding-Tube, and In-Hospital Postoperative Mortality in Children: a retrospective cohort study. Anesth Analg. 2023;136:1133-42.
12. Canarie MF, Barry S, Carroll CL, Hassinger A, Kandil S, Li S, et al. Risk factors for delayed enteral nutrition in critically Ill children. Pediatr Crit Care Med. 2015;16:e283-9.
13. Meier JD, Valentine KJ, Hagedorn C, Hartling C, Gershan W, Muntz HR, et al. Emergency department use among children with tracheostomies: avoidable visits. J Pediatr Rehabil Med. 2015;8:105-11.
14. Nageswaran S, Golden SL, Gower WA, King NMP. Caregiver perceptions about their decision to pursue tracheostomy for children with medical complexity. J Pediatr. 2018;203:354-60.e1.
15. Watters KF. Tracheostomy in infants and children. Respir Care. 2017;62:799-25.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.42027
Eurasian J Emerg Med. 2025;24(3): 219-27

Emergency Department Evaluation of NLR, PLR, and LMR for Predicting Contrast-Induced Nephropathy in Acute Coronary Syndrome Patients: A Retrospective Observational Study

Veysi Siber[1], Serdal Ateş[2], Kaan Yıldız[1], Merve Berfin Önen[1], Meryem Sara Dayan[1], Burak Katipoğlu[1]

[1]University of Health Sciences Türkiye, Ankara Etlik City Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
[2]University of Health Sciences Türkiye, Ankara Training and Research Hospital, Clinic of Emergency Medicine, Ankara, Türkiye

Abstract

Aim: Contrast-induced nephropathy (CIN) is a significant complication following contrast exposure, especially in patients with acute coronary syndrome (ACS). Inflammation plays a critical role in CIN pathogenesis. This study aimed to evaluate the prognostic value of hematologic inflammation indices-neutrophil-to-lymphocyte ratio (NLR), lymphocyte-to-monocyte ratio (LMR), platelet-to-lymphocyte ratio (PLR) and the lactate-to-albumin ratio (LAR) in predicting CIN development in emergency department patients diagnosed with ACS.

Materials and Methods: This retrospective, single-center study included ACS patients who underwent contrast-enhanced imaging or percutaneous coronary intervention between July and December 2024. Inflammatory indices were calculated from complete blood count on admission and biochemical parameters. Patients were categorized as CIN (+) or CIN (–) based on Kidney Disease Improving Global Outcomes criteria. ROC curve analysis was used to assess diagnostic performance.

Results: Among 814 patients, CIN developed in 89 (10.9%). Compared to the non-CIN group, patients with CIN were older (71 vs. 61 years, $p<0.001$) and had significantly higher NLR (3.3 vs. 2.6, $p<0.001$), PLR (143 vs. 117, $p=0.006$), and lower LMR (2.8 vs. 3.3, $p=0.016$). ROC analysis showed that age had the highest area under the curve (0.697), followed by NLR (0.615), PLR (0.590), and LMR (0.578). LMR showed the highest sensitivity (86%), while NLR had the highest specificity (84%). LAR was not significantly associated with CIN ($p=0.208$).

Conclusion: Hematologic inflammation indices such as NLR, LMR, and PLR may serve as cost-effective, accessible tools for early CIN risk stratification in ACS patients. These biomarkers may complement established clinical risk scores and assist in guiding nephroprotective strategies in the emergency setting.

Keywords: Acute coronary syndrome, acute kidney injury, neutrophils, lymphocytes, monocytes, platelets

Introduction

Acute coronary syndrome (ACS) is one of the leading causes of cardiovascular mortality worldwide and represents a critical clinical condition frequently encountered in emergency departments, where diagnosis and treatment are time-sensitive. To confirm the diagnosis of ACS and evaluate concomitant coronary artery disease, imaging modalities involving contrast media are often used. However, the use of intravenous contrast agents, particularly in patients with comorbidities, increases the risk of developing contrast-induced nephropathy (CIN), which may result in prolonged hospital stay, increased healthcare costs, and worsened clinical outcomes (1).

Although several clinical scoring systems have been proposed to predict the development of CIN, many of these tools are complex and time-sensitive, limiting their applicability in emergency settings. As a result, there is an increasing need for practical, rapid, cost-effective, and objective biomarkers to support timely clinical decision-making. In this context, hematologic inflammatory parameters derived from complete blood count (CBC) data have gained prominence. Notably, indices such as the neutrophil-to-lymphocyte ratio (NLR), lymphocyte-to-monocyte ratio (LMR), and

Corresponding Author: Veysi Siber MD, University of Health Sciences Türkiye, Ankara Etlik City Hospital, Clinic of Emergency Medicine, Ankara, Türkiye
E-mail: veysiber.ss@gmail.com **ORCID ID:** orcid.org/0000-0003-2856-8303

Cite this article as: Siber V, Ateş S, Yıldız K, Önen MB, Dayan MS, Katipoğlu B. Emergency department evaluation of NLR, PLR, and LMR for predicting contrast-induced nephropathy in acute coronary syndrome patients: a retrospective observational study. Eurasian J Emerg Med. 2025;24(3): 219-27.

Received: 01.08.2025
Accepted: 18.08.2025
Published: 10.09.2025

platelet-to-lymphocyte ratio (PLR) have demonstrated prognostic value as surrogate markers of systemic inflammation in various cardiovascular, neoplastic, and infectious conditions (2-4).

NLR increases in response to acute inflammatory processes due to neutrophilia and lymphocyte suppression, making it a reliable marker of inflammation severity. Elevated NLR has been significantly associated with the development of CIN in patients diagnosed with ACS undergoing percutaneous interventions, and has been identified as an independent predictor in previous studies (5). Similarly, LMR reflects inflammatory activity through increased monocyte levels and decreased lymphocyte counts, both of which play pivotal roles in the regulation of immune responses. Low LMR has been reported to be associated with an increased risk of CIN (6). On the other hand, PLR serves as a marker that reflects both the pro-inflammatory role of platelets (PLTs) and the immunomodulatory function of lymphocytes. A growing body of literature supports the correlation between elevated PLR and adverse cardiovascular outcomes, as well as renal dysfunction, highlighting its potential prognostic utility (7,8).

Beyond hematologic indices, we also evaluated the lactate-to-albumin ratio (LAR). This parameter combines two routinely measured laboratory values: lactate, which reflects tissue hypoperfusion and metabolic stress, and albumin, which indicates nutritional and inflammatory status as a negative acute-phase reactant. Considering that both markers are commonly associated with systemic illness and organ dysfunction, we hypothesized that LAR might also serve as a potential predictor of CIN in patients with ACS.

In conclusion, hematologic inflammatory indices such as NLR, LMR, and PLR, which are easily calculated, cost-free, and widely accessible, may serve as practical tools for predicting CIN in patients presenting to the emergency department with ACS and undergoing contrast-enhanced procedures. Evaluating the predictive utility of these biomarkers may facilitate more effective risk stratification in clinical practice. The findings of this study may contribute to the development of more informed and preventive strategies, particularly in high-risk individuals. Accordingly, the aim of this study is to investigate the prognostic value of hematologic inflammatory markers, including NLR, LMR, PLR, and the LAR, in predicting the development of CIN in ACS patients exposed to contrast media.

Materials and Methods

Study Design and Setting

This manuscript was prepared in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology guidelines to ensure transparent and comprehensive reporting. The study was designed as a single-center, retrospective, descriptive and analytical observational study. Ethical approval was obtained from the Clinical Research Ethics Committee of University of Health Sciences Türkiye, Ankara Etlik City Hospital (decision number: AEŞH-BADEK1-2025-174, date: 07.05.2025). The study was conducted in accordance with the principles of the Declaration of Helsinki. Since all participants were anonymized, informed consent was not required.

Study Population and Definitions

The target population of this study consisted of patients who presented to the emergency department with suspected ACS between July 1, 2024, and December 31, 2024, whose diagnosis was confirmed, and who underwent diagnostic or interventional imaging with contrast media.

ACS is a cardiovascular emergency that requires prompt diagnosis and treatment, and encompasses unstable angina pectoris, ST-elevation myocardial infarction (STEMI), and non–STEMI, all of which result from myocardial ischemia (9). The diagnosis of ACS was based on clinical symptoms, electrocardiogram (ECG) changes, and elevated cardiac biomarkers.

CIN was defined as an absolute increase in serum creatinine of ≥0.5 mg/dL or a relative increase of ≥25% from baseline within 72 hours after contrast administration (10).

Inclusion Criteria

- Age ≥18 years
- Confirmed diagnosis of ACS based on clinical presentation, ECG findings, or elevated troponin levels
- Undergoing a diagnostic or interventional procedure involving contrast media
- Availability of serum creatinine measurements at presentation and within 72 hours after contrast exposure

Exclusion Criteria

- Presence of active infection or clinical signs of sepsis
- History of chronic inflammatory or autoimmune disease
- Patients with acute or chronic liver disease
- Patients undergoing chronic medical treatment with steroids or non-steroidal anti-inflammatory drugs, or with a history of organ transplantation
- Use of systemic corticosteroids within the past 2 weeks
- Exposure to contrast media within the past 2 weeks

- History of infection within the past 2 weeks (e.g., respiratory tract infection, urinary tract infection, dental abscess, etc.)
- History of hematologic malignancy or immunosuppressive therapy
- Patients undergoing emergency surgical procedures
- Presence of pre-existing renal failure prior to contrast administration (eGFR <30 mL/min/1.73 m^2)
- Patients who received nephroprotective agents such as N-acetylcysteine prior to contrast exposure
- Cases with missing laboratory data
- Patients with multiple exposures to contrast media within 24 hours
- Patients without serum creatinine re-evaluation within 72 hours after contrast administration
- Pregnant patients

Data Collection

Within the specified study period, patients presenting to the emergency department were screened using ICD-10 codes to identify those diagnosed with ACS. Among these, patients that underwent contrast-enhanced imaging procedures were selected. Data were retrospectively retrieved from the hospital information management system. Variables evaluated in the study included demographic, clinical, hematologic, and biochemical parameters. Demographic data included age and sex. Clinical variables comprised heart rate, presence of STEMI, anemia, diabetes mellitus (DM), hypertension (HT), and prior history of percutaneous coronary intervention (PCI). Cardiac rhythm was documented using a standard 12-lead ECG.

In the hematologic evaluation, CBC data were used to assess hemoglobin (HGB), PLT count, mean PLT volume, neutrophil, lymphocyte, and monocyte levels. From these parameters, LMR, NLR, and PLR were calculated.

For the biochemical evaluation, serum glucose, creatinine, albumin, and lactate levels were recorded, and LAR was calculated accordingly. Venous blood samples were obtained within 10 minutes of presentation to the emergency department and prior to coronary angiography. Hematologic parameters, including monocyte, lymphocyte, and neutrophil counts, and HGB, were analyzed within 30 minutes of sampling using an automated hematology analyzer. Biochemical analyses-including creatinine, lactate, albumin, and glucose levels-were performed using standard laboratory techniques. These laboratory results were documented for all patients. Serum creatinine was re-measured in all patients within 72 hours following contrast exposure.

CIN was evaluated according to the diagnostic criteria defined by the Kidney Disease Improving Global Outcomes guidelines. Accordingly, patients who demonstrated an absolute increase in serum creatinine of ≥0.5 mg/dL, or a relative increase of ≥25% from baseline within 72 hours following contrast administration, were classified as CIN-positive; those who did not meet these thresholds were categorized as CIN-negative.

To ensure patient confidentiality, all data were processed in a de-identified format. A unique code was assigned to each individual, and personally identifiable information was excluded from the analysis. Missing data were checked on a variable-by-variable basis. Patients with any missing clinical or laboratory parameters were excluded from the study in accordance with the predefined exclusion criteria.

To ensure data accuracy, parallel data entry was performed independently by two researchers. A random 10% sample was selected for cross-checking. Inconsistent entries were verified and corrected through re-examination of the hospital database.

The data collection process was coordinated by the principal investigator (V.S.) and all clinical, laboratory, and diagnostic variables were systematically compiled according to the predefined data collection protocol.

Statistical Analysis

All statistical analyses were performed using SPSS version 27.0 (IBM Corp., Armonk, NY, USA) and Jamovi version 2.5.7 software. Descriptive statistics for categorical variables were presented as frequencies and percentages, while continuous variables were expressed as medians with [interquartile ranges (IQR), 25^{th}-75^{th} percentiles)]. The normality of distribution for continuous variables was assessed using the Kolmogorov–Smirnov test and histogram analysis.

Comparisons of categorical variables were conducted using the chi-square test. For continuous variables with normal distribution, the Student's t-test and Welch's t-test were used depending on the homogeneity of variance, which was evaluated by Levene's test. The Mann–Whitney U test was employed for comparing non-normally distributed continuous variables.

To assess the predictive performance of inflammatory indices for CIN, ROC curve analysis was conducted. From this analysis, threshold values, sensitivity, specificity, positive predictive value, negative predictive value (NPV), positive likelihood ratio, and negative likelihood ratio were calculated. The Youden index was used to determine the optimal cut-off points.

The DeLong test was applied to compare the diagnostic performance [area under the curve, (AUC)] of different biomarkers. A p value of <0.05 was considered statistically significant.

Results

A total of 1.326 patients were initially identified based on ICD-10 codes. Of these, 398 patients were excluded due to missing data. The remaining 928 patients were evaluated for eligibility. Subsequently, patients were excluded for the following reasons: history of contrast media exposure within the past two weeks (n=6); liver disease (n=8); pre-existing renal failure prior to contrast administration (n=13); lack of follow-up creatinine measurement within 72 hours (n=57); multiple exposures to contrast media within 24 hours (n=11); active infection or a history of infection within the past two weeks (n=19). After applying these exclusion criteria, a total of 814 patients were included in the final analysis (Figure 1).

The median age of the participants was 62 years (IQR: 53-71), and 72.4% were male. The most common comorbidity was HT, observed in 52.9% of patients, followed by DM (38.1%), history of PCI (33.7%) and anemia (7.9%). CIN developed in 10.9% of patients. Additionally, 16.5% of the study population was diagnosed with STEMI (Table 1).

The median age of patients who developed CIN was significantly higher than those who did not (p<0.001). HT was more prevalent among patients with CIN (74.2%) than those without CIN (50.3%)

Table 1. Distribution of patients' demographic characteristics, comorbidities, laboratory parameters, and clinical features

Age, years, median (IQR)	62 (53-71)
Sex, n (%) Male Female	 589 (72.4%) 225 (27.6%)
Hypertension, n (%) Diabetes mellitus Anemia History of PCI	431 (52.9%) 310 (38.1%) 64 (7.9%) 274 (33.7%)
Heart rate, median (IQR)	76 (66-89)
Laboratory parameters, median (IQR) Hemoglobin, g/dL Platelet, 109/L Neutrophil, 109/L Lymphocyte, 109/L Monocyte, 109/L Mean platelet volume, fL Blood glucose, mg/dL Creatinine, mg/dL Post-anjio creatinine, mg/dL Lactate, mmol/L Albumin, g/L Lactate / albumin ratio LMR NLR PLR	 14.4 (12.9-15.5) 255 (214-300) 5.8 (4.5-7.8) 2.1 (1.5-2.8) 0.67 (0.52-0.83) 10.3 (9.7-11.0) 130 (104-182) 0.94 (0.81-1.11) 0.93 (0.79-1.10) 1.7 (1.3-2.4) 40 (38-43) 4.3 (3.3-5.8) 3.3 (2.4-4.4) 2.7 (1.8-4.3) 119 (89-163)
CIN, n (%)	89 (10.9%)
STEMI, n (%)	134 (16.5%)

IQR: Interquartile range, PCI: Percutaneous coronary intervention, NLR: Neutrophil-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, LMR: Lymphocyte-to-monocyte ratio, CIN: Contrast-induced nephropathy, STEMI: ST-elevation myocardial infarction

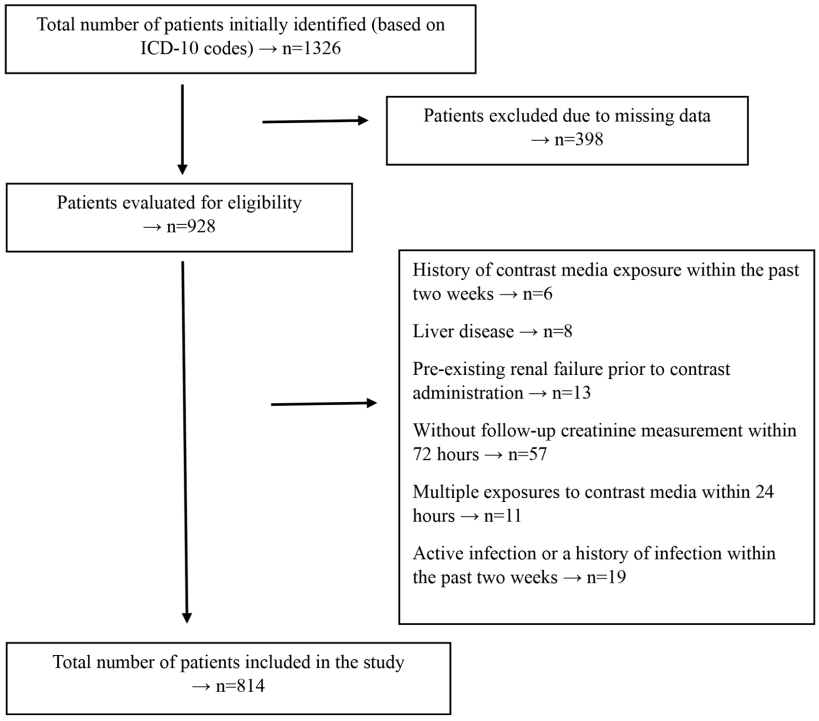


Figure 1. Patient flow diagram

(p<0.001). Additionally, a prior history of PCI was significantly more common in the CIN group (p=0.017) (Table 2).

Median heart rate was higher in patients with CIN (p=0.020). HGB levels were significantly lower in the CIN group (p=0.010), while PLT was lower (p=0.029) (Table 2).

Patients with CIN had significantly lower lymphocyte counts (p<0.001) and higher blood glucose levels (p=0.014). Both baseline and post-procedural creatinine levels were significantly elevated in the CIN group (Table 2).

The albumin level was slightly lower in the CIN group (p=0.019). Moreover, LMR, NLR, and PLR were significantly higher in patients who developed CIN (LMR: p=0.016; NLR: p<0.001; PLR: p=0.006) (Table 2).

Diagnostic Accuracy of Age, NLR, PLR, and LMR in Predicting CIN

As shown in Table 3, the cut-off values determined for predicting CIN were >64 for age, >5.2 for NLR, >137 for PLR, and <2.0 for LMR. Among these, age had the highest AUC: 0.697, followed by NLR (AUC: 0.615), PLR (AUC: 0.590), and LMR (AUC: 0.578) (Figure 2).

Notably, LMR demonstrated the highest sensitivity [(86%; 95% confidence interval (CI): 83-88)], although its specificity remained low (29%; 95% CI: 20-40). In contrast, NLR showed the highest specificity (84%; 95% CI: 81-86) with relatively low sensitivity (36%; 95% CI: 26-47). NPV were consistently high across all markers, ranging from 0.91 to 0.95 (Table 3).

Pairwise comparisons using the DeLong test revealed that the predictive accuracy of age was significantly superior to that of NLR (AUC difference = 0.082; 95% CI: 0.001-0.163; p=0.046), PLR (AUC difference = 0.107; 95% CI: 0.031-0.183; p=0.006), and LMR (AUC difference = 0.119; 95% CI: 0.044-0.194; p=0.002) (Table 3).

These results indicate that while age exhibited the best overall discriminatory performance, LMR, and NLR may also serve as supplementary inflammatory markers in predicting the risk of CIN.

Table 2. Association of patients' demographic characteristics, comorbidities, laboratory parameters, and clinical features with the development of CIN

	CIN+ (n=89)	CIN- (n=725)	p value
Age, years, median (IQR)	71 (64-76)	61 (52-70)	<0.0011***
Sex, n (%) Male Female	 58 (65.2%) 31 (34.8%)	 58 (65.2%) 31 (34.8%)	 0.1082
Comorbidities, n (%) Hypertension, n (%) Diabetes mellitus Anemia History of PCI	 66 (74.2%) 37 (41.6%) 8 (9.0%) 40 (44.9%)	 365 (50.3%) 273 (37.7%) 56 (7.7%) 234 (32.3%)	 <0.0012*** 0.4732 0.6762 0.0172*
Heart rate, median (IQR)	81 (70-93)	75 (66-88)	0.0203*
Laboratory parameters, median (IQR) Hemoglobin, g/dL Platelet, 109/L Neutrophil, 109/L Lymphocyte, 109/L Monocyte, 109/L Mean platelet volume, fL Blood glucose, mg/dL Creatinine, mg/dL Post-anjio creatinine, mg/dL Lactate, mmol/L Albumin, g/L Lactate / albumin ratio LMR NLR PLR	 13.4 (11.7-15.2) 234 (194-293) 6.1 (4.8-8.4) 1.7 (1.2-2.6) 0.67 (0.51-0.79) 10.4 (9.8-11.0) 152 (114-210) 1.07 (0.94-1.40) 1.43 (1.21-1.87) 1.7 (1.4-2.5) 41 (38-42) 4.3 (3.5-6.6) 2.8 (1.8-4.1) 3.3 (2.0-6.9) 143 (101-203)	 14.4 (13.1-15.6) 257 (215-302) 5.8 (4.5-7.7) 2.1 (1.6-2.9) 0.67 (0.52-0.84) 10.3 (9.7-11.0) 129 (103-181) 0.93 (0.80-1.09) 0.91 (0.78-1.05) 1.7 (1.3-2.3) 41 (38-43) 4.3 (3.3-5.6) 3.3 (2.5-4.5) 2.6 (1.7-4.1) 117 (88-159)	 0.0101* 0.0291* 0.0873 <0.0013*** 0.2703 0.2891 0.0143* <0.0013*** <0.0013*** 0.4153 0.0193* 0.2083 0.0163* <0.0013*** 0.0063**
STEMI, n (%)	14 (15.7 %)	120 (16.6 %)	0.8442

Statistical comparisons were made using the independent samples t-test', chi-square test', and Mann–Whitney U test' as appropriate. ***p<0.001, **p<0.01, *p<0.05
IQR: Interquartile range, PCI: Percutaneous coronary intervention, CIN: Contrast-induced nephropathy, NLR: Neutrophil-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, LMR: Lymphocyte-to-monocyte ratio, STEMI: ST-elevation myocardial infarction

Table 3. Cut-off values for age, NLR, PLR and LMR in predicting CIN

	Cut-off	AUC (95% CI)	SEN (95% CI)	SPE (95% CI)	PPD (95% CI)	NPD (95% CI)	+LR (95% CI)	-LR (95% CI)
Age	**>64**	0.697 (0.642-0.753)	0.78 (0.67-0.86)	0.58 (0.55-0.62)	0.19 (0.17-0.21)	0.95 (0.93-0.97)	1.9 (1.6-2.1)	0.4 (0.3-0.6)
NLR	>5.2	0.615 (0.550-0.680)	0.36 (0.26-0.47)	0.84 (0.81-0.86)	0.21 (0.16-0.27)	0.91 (0.90-0.93)	2.2 (1.6-3.1)	0.8 (0.6-0.9)
PLR	>137	0.590 (0.523-0.657)	0.55 (0.44-0.66)	0.65 (0.60-0.67)	0.16 (0.13-0.19)	0.92 (0.90-0.94)	1.5 (1.2-1.9)	0.7 (0.5-0.9)
LMR	<2.0	0.578 (0.513-0.643	0.86 (0.83-0.88)	0.29 (0.20-0.40)	0.91 (0.90-0.92)	0.20 (0.15-0.27)	1.2 (1.1-1.4)	0.5 (0.3-0.7)

CIN: Contrast-induced nephropathy, NLR: Neutrophil-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, LMR: Lymphocyte-to-monocyte ratio, AUC: Area under the curve, PPV: Positive predictive value, NPV: Negative predictive value, SEN: Sensitivity, SPE: Specificity, +LR: Positive likelihood ratio, -LR:Negative likelihood ratio, NPD: New predictive determinant, PPD: Pre-procedural data, CI: Confidence interval
Delong test pairwise comparisons results
Age vs NLR: AUC Difference =0.082 (%95 Cl: 0.001, 0.163), p=0.046
Age vs PLR: AUC Difference =0.107 (%95 Cl: 0.031, 0.183), p=0.006
NLR vs PLR: AUC Difference =0.025 (%95 Cl: -0.020, 0.069) p=0.278
Age vs LMR: AUC Difference =0.119 (%95 Cl: 0.044, 0.194) p=0.002
NLR vs LMR: AUC Difference =0.037 (%95 Cl: -0.015, 0.088) p=0.165

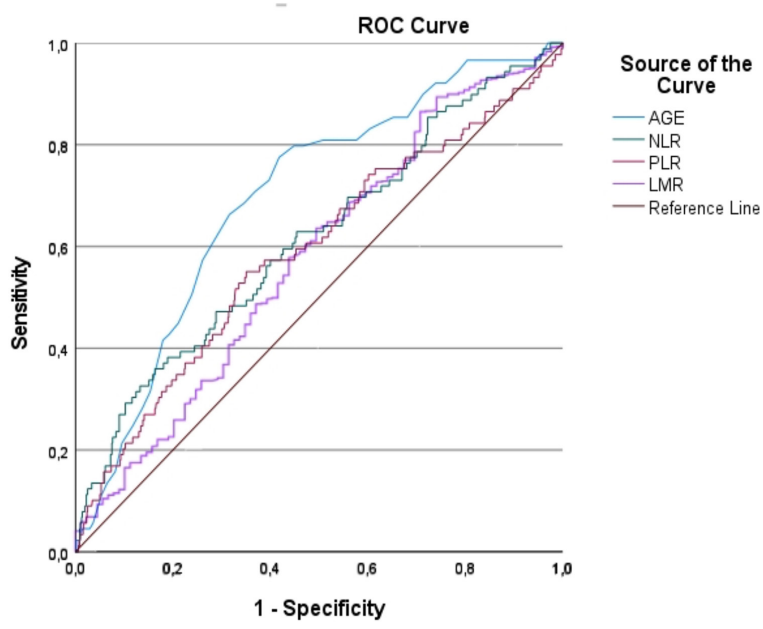


Figure 2. ROC curves of age, NLR, PLR and LMR in predicting CIN

ROC: Receiver operating characteristic, NLR: Neutrophil-to-lymphocyte ratio, LMR: Lymphocyte-to-monocyte ratio, PLR: Platelet-to-lymphocyte ratio, CIN: Contrast-induced nephropathy

Discussion

In this retrospective observational study, we evaluated the prognostic utility of NLR, LMR, PLR, and LAR in predicting the development of CIN in patients diagnosed with ACS and exposed to contrast media in the emergency department setting. According to our findings, elevated NLR and PLR levels, as well as decreased LMR values, were significantly associated with the development of CIN. Moreover, among all evaluated parameters, age was identified as the strongest independent predictor.

CIN is a significant iatrogenic complication, particularly among individuals with cardiovascular comorbidities. It is believed to develop through mechanisms such as impaired renal perfusion, inflammation, and oxidative stress (11,12). This underlying pathophysiology supports the rationale for using inflammatory biomarkers as potential predictors of CIN.

In this context, NLR, which can be easily calculated from routine CBC parameters, has emerged as a widely used and practical marker of systemic inflammation in clinical practice. Previous studies have reported that NLR is independently associated with the development of CIN in patients who experience myocardial infarction and undergo PCI (3,13,14). Consistent with these findings, our study also demonstrated a significant association between elevated NLR levels (>5.2) and CIN development (13,15).

Notably, the high specificity (84%) and NPV (91%) observed in our analysis suggest that patients with low NLR values are at a considerably lower risk for CIN.

LMR reflects both the cellular and immunologic aspects of the inflammatory response by indicating monocyte elevation and lymphocyte suppression. Previous evidence has shown that low LMR levels are associated with various clinical conditions, including sepsis, cardiovascular disease, and chronic kidney disease (6,16,17). In our study, an LMR threshold of <2.0 demonstrated high sensitivity (86%) for predicting CIN, suggesting that LMR may serve as a valuable screening tool for identifying patients at high risk for CIN.

PLR serves as a biomarker positioned at the intersection of inflammation and thrombosis, as it reflects both the pro-inflammatory activity of PLTs and the immunoregulatory function of lymphocytes (18). In our study, a PLR value greater than 137 was found to be significantly associated with the development of CIN. This finding is consistent with previous literature demonstrating the association of elevated PLR with cardiovascular mortality and renal injury (19-21).

LAR, which has been increasingly used in the literature as a prognostic biomarker for mortality in sepsis and critically ill patients (22-24), was not found to be significantly associated with CIN development in our study. This discrepancy may be attributed to the relatively lower degree of systemic inflammation in ACS patients compared to more severe inflammatory conditions such as sepsis.

Another important factor influencing the development of CIN is the variability in contrast media exposure between diagnostic and interventional coronary procedures. While diagnostic coronary angiography usually involves relatively low volumes, interventional procedures such as PCI typically require substantially higher amounts, particularly in complex or multi-vessel interventions. Since our study did not include standardized data on the type or amount of contrast administered, this procedural heterogeneity may have contributed to CIN risk and should be acknowledged as a limitation.

According to the ROC analysis, age demonstrated the highest AUC: 0.697 when compared to hematologic indices such as NLR, LMR, and PLR. The association between advanced age and increased CIN risk may be explained by age-related physiological changes, including a decline in glomerular filtration rate, heightened susceptibility to oxidative damage, and endothelial dysfunction-all of which contribute to age as an independent and significant risk factor for CIN (25,26).

Although the AUC values of NLR, PLR, and LMR reached statistical significance, their relatively low magnitude restricts their discriminatory power. Therefore, these indices alone are unlikely to substantially influence clinical decision-making. Rather, they should be interpreted as adjunctive markers, supporting established risk scores such as the Mehran score, to improve the overall accuracy of CIN risk prediction.

Our findings suggest that hematologic parameters such as NLR, LMR, and PLR may contribute to CIN risk stratification in clinical practice. These biomarkers offer several advantages, including wide availability, low cost, repeatability, and ease of calculation. However, relying solely on these indices for clinical decision-making is insufficient. Instead, they should be used as complementary tools alongside established risk prediction models, such as the Mehran score, to enhance the accuracy of CIN risk assessment (11).

Nevertheless, the retrospective and single-center nature of our study restricts external validity, and the findings should be interpreted with caution when applied to different populations or healthcare settings.

Study Limitations

Although this study possesses several strengths, certain limitations should be acknowledged. First, its retrospective and single-center design restricts the ability to establish causal relationships and limits the generalizability of the findings to wider populations. Second, hematologic inflammatory parameters were measured only once at the time of emergency department admission, whereas inflammation is a dynamic process; therefore, the lack of serial measurements may have reduced the prognostic value of these biomarkers. Third, despite excluding patients with active infection, chronic inflammatory disease, or recent corticosteroid use, the possibility of subclinical inflammation or other systemic processes could not be entirely ruled out. Finally, details regarding nephroprotective interventions, hydration protocols, contrast agent type, and dose, and other procedural factors were not standardized or fully recorded, and these uncontrolled variables may have acted as potential confounders influencing CIN development. Similarly, the type and volume of contrast media varied across diagnostic and interventional procedures, with the latter typically requiring substantially higher contrast doses. This lack of standardization limited our ability to assess the impact of contrast exposure in detail and may have influenced the incidence of CIN.

Conclusion

In conclusion, hematologic inflammation indices such as NLR, LMR, and PLR may serve as useful adjuncts in predicting CIN in patients with ACS. While age emerged as the strongest independent predictor, these indices may contribute to early risk stratification and help guide the implementation of nephroprotective strategies in clinical practice.

Ethics

Ethics Committee Approval: Ethical approval was obtained from the Clinical Research Ethics Committee of Ankara Etlik City Hospital (decision number: AEŞH-BADEK1-2025-174, date: 07.05.2025).

Informed Consent: Retrospective study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: V.S., Concept: V.S., B.K., Design: V.S., S.A., B.K., Data Collection or Processing: V.S., K.Y., M.B.Ö., M.S.D., Analysis or Interpretation: V.S., S.A., B.K., Literature Search: K.Y., M.B.Ö., M.S.D., Writing: V.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. Yang Y, George KC, Luo R, Cheng Y, Shang W, Ge S, et al. Contrast-induced acute kidney injury and adverse clinical outcomes risk in acute coronary syndrome patients undergoing percutaneous coronary intervention: a meta-analysis. BMC Nephrol. 2018;19:374.
2. Elcik D, Yarlioglu M, Oksuz F, Duran M, Inanc MT. Relationship between lymphocyte to monocyte ratio and contrast-induced nephropathy in acute coronary syndrome patients treated with percutaneous coronary intervention. J Med Res Rev. 2024;2:110-9.
3. Tanık VO, Çınar T, Velibey Y, Öz A, Kalenderoğlu K, Gümüşdağ A, et al. Neutrophil-to-lymphocyte ratio predicts contrast-induced acute kidney injury in patients with st-elevation myocardial infarction treated with primary percutaneous coronary intervention. J Tehran Heart Cent. 2019;14:59-66.
4. Jiang J, Ji HY, Xie WM, Ran LS, Chen YS, Zhang CT, et al. Could platelet-to-lymphocyte ratio be a predictor for contrast-induced nephropathy in patients with acute coronary syndrome?: a systematic review and meta-analysis. Medicine (Baltimore). 2019;98:e16801.
5. Yang Z, Qiao Y, Wang D, Yan G, Tang C. Association between inflammatory biomarkers and contrast-induced acute kidney injury in ACS patients undergoing percutaneous coronary intervention: a cross-sectional study. Angiology. 2024;75:831-40.
6. Karauzum I, Karauzum K, Acar B, Hanci K, Bildirici HIU, Kilic T, et al. Predictive value of lymphocyte-to-monocyte ratio in patients with contrast-induced nephropathy after percutaneous coronary intervention for acute coronary syndrome. J Transl Int Med. 2021;9:123-30.
7. Butt K, D'Souza J, Yuan C, Jayakumaran J, Nguyen M, Butt HI, et al. Correlation of the neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR) with contrast-induced nephropathy in patients with acute coronary syndrome undergoing percutaneous coronary interventions. Cureus. 2020;12:e11879.
8. Zorlu C, Koseoglu C. Comparison of the relationship between inflammatory markers and contrast-induced nephropathy in patients with acute coronary syndrome after coronary angiography. Angiology. 2020;71:249-55.
9. Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA, et al. Fourth universal definition of myocardial infarction (2018). J Am Coll Cardiol. 2018;72:2231-64.
10. Silvain J, Collet JP, Montalescot G. Contrast-induced nephropathy: the sin of primary percutaneous coronary intervention? Eur Heart J. 2014;35:1504-6.
11. Mehran R, Aymong ED, Nikolsky E, Lasic Z, Iakovou I, Fahy M, et al. A simple risk score for prediction of contrast-induced nephropathy after percutaneous coronary intervention: development and initial validation. J Am Coll Cardiol. 2004;44:1393-9.
12. McCullough PA, Adam A, Becker CR, Davidson C, Lameire N, Stacul F, et al. Epidemiology and prognostic implications of contrast-induced nephropathy. Am J Cardiol. 2006;98:5K-13K.
13. Kaya A, Kaya Y, Topçu S, Günaydin ZY, Kurt M, Tanboğa IH, et al. Neutrophil-to-lymphocyte ratio predicts contrast-induced nephropathy in patients undergoing primary percutaneous coronary intervention. Angiology. 2014;65:51-6.
14. Tamhane UU, Aneja S, Montgomery D, Rogers EK, Eagle KA, Gurm HS. Association between admission neutrophil to lymphocyte ratio and outcomes in patients with acute coronary syndrome. Am J Cardiol. 2008;102:653-7.
15. Dönmez E, Özcan S, Şahin İ, Okuyan E. The association between hemogram parameters and the development of contrast-induced nephropathy in patients presenting with non-ST-elevation myocardial infarction. Adv Lab Med. 2023;4:308-13.
16. Garcia-Rojo D, Prera A, Muñoz-Rodriguez J, Oliva JC, Dominguez A, Prats J. Prognostic value of lymphocyte-to-monocyte ratio previously determined to surgery in patients with non-metastatic renal cell carcinoma: a systematic review and a PRISMA-compliant meta-analysis. Medicine (Baltimore). 2021;100:e24152.
17. Hu X, Qin X, Gu X, Wang H, Zhou W. Effect of lymphocyte-to-monocyte ratio on survival in septic patients: an observational cohort study. Arch Med Sci. 2020;20:790-7.
18. Azab B, Shah N, Akerman M, McGinn JT Jr. Value of platelet/lymphocyte ratio as a predictor of all-cause mortality after non-ST-elevation myocardial infarction. J Thromb Thrombolysis. 2012;34:326-34.
19. Demircelik MB, Kurtul A, Ocek H, Cakmak M, Ureyen C, Eryonucu B. Association between platelet-to-lymphocyte ratio and contrast-induced nephropathy in patients undergoing percutaneous coronary intervention for acute coronary syndrome. Cardiorenal Med. 2015;5:96-104.
20. Gasparyan AY, Ayvazyan L, Mukanova U, Yessirkepov M, Kitas GD. The platelet-to-lymphocyte ratio as an inflammatory marker in rheumatic diseases. Ann Lab Med. 2019;39:345-57.
21. Canpolat U, Aydogdu S, Turak O, Aras D, Tok D, Ozcan F, et al. Association of platelet-to-lymphocyte ratio with inflammation and fibrosis in nonalcoholic fatty liver disease. Can J Gastroenterol Hepatol. 2017;2017:1-8.
22. Wang Y, Yu H. Association between lactate to albumin ratio and mortality among sepsis associated acute kidney injury patients. BMC Infect Dis. 2025;25:414.
23. Bou Chebl R, Jamali S, Sabra M, Safa R, Berbari I, Shami A, et al. Lactate/albumin ratio as a predictor of in-hospital mortality in septic patients presenting to the emergency department. Front Med (Lausanne). 2020;7:550182.

24. Gharipour A, Razavi R, Gharipour M, Mukasa D. Lactate/albumin ratio: an early prognostic marker in critically ill patients. Am J Emerg Med. 2020;38:2088-95.

25. Khalfallah M, Allaithy A, Maria DA. Incidence, predictors and outcomes of contrast induced nephropathy in patients with ST elevation myocardial infarction undergoing primary percutaneous coronary intervention. Glob Heart. 2021;16:57.

26. Weisbord SD, Mor MK, Resnick AL, Hartwig KC, Sonel AF, Fine MJ, et al. Prevention, incidence, and outcomes of contrast-induced acute kidney injury. Arch Intern Med. 2008;168:1325-32.

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Original Article
DOI: 10.4274/eajem.galenos.2025.32748
Eurasian J Emerg Med. 2025;24(3): 228-31

Clinical Profile and Prognostic Indicators of Papaver Rhoeas Poisoning in Adults: A Retrospective Analysis from Türkiye

Emin Fatih Vişneci

University of Health Sciences Türkiye, Konya City Hospital, Clinic of Emergency Medicine, Konya, Türkiye

Abstract

Aim: The corn poppy is often consumed as a herbal tea, believed to have various health benefits in traditional medicine; it is especially commonly consumed as food in Anatolia. The aim of this study is to assess the clinical features of patients who presented to the emergency department due to corn poppy poisoning, evaluate the duration of admission, lactate levels, and length of hospital stay, and investigate their effects on prognosis.

Materials and Methods: This retrospective study included 15 patients divided into two groups: mildly symptomatic (n=8) and severely symptomatic (n=7). Gastrointestinal, neurological, and cardiac symptoms, latency time between ingestion and symptom onset, lactate levels, and hospitalization duration were compared. Inferential statistical analyses were performed using Fisher's exact test for categorical variables, such as time to presentation category and hospital stay length category, due to the small sample size. The Mann-Whitney U test was applied for comparisons of symptom scores between groups. A p value of <0.05 was considered statistically significant.

Results: Severe cases exhibited frequent gastrointestinal and both mild and severe neurological symptoms, but no cardiac manifestations. In contrast, mild cases presented earlier, had lower lactate levels, and shorter hospital stays. Severe neurological symptoms and elevated lactate levels were significantly associated with poor prognosis.

Conclusion: Corn poppy poisoning can present emergency services. This study also suggests that toxicity severity may not solely depend on the quantity ingested but also on the plant's developmental stage. Public and healthcare providers' awareness of the potential toxicity of traditionally used plants like Papaver rhoeas should be increased.

Keywords: Papaver rosea, poisoning, traditional medicine, emergency

Introduction

Papaver rhoeas, commonly known as corn poppy, is widely cultivated in Türkiye and used in folk medicine for ailments such as cough, insomnia, and gastric discomfort (1,2). In addition, depending on the climate conditions, poppies emerge in March or April and bloom within 1-2 months. (Figure 1.)

The use of plants for healing purposes is as common as the consumption of fresh herbs as food, especially in the spring months in Anatolia (3).

Despite its perceived harmlessness due to its herbal origin, the plant contains alkaloid compounds capable of producing toxic effects, particularly on the central nervous and gastrointestinal systems. Cardiac involvement has also been reported in some cases (4,5). Alkaloid components, when taken in excessive doses, can cause depression, especially with a sedative effect on the central nervous system; this may manifest as nausea, vomiting, changes in consciousness, convulsions, and miotic pupils (6). In addition, it has been reported in emergency department cases that this plant can produce anticholinergic and antidopaminergic effects, resulting in symptoms similar to acute anticholinergic syndrome (such as irritation, agitation, hypertension, tachycardia, muscle twitching) (7).

The increasing popularity of herbal products and the general assumption that "natural" means safe may contribute to their misuse, particularly in rural populations. Existing literature

Corresponding Author: Emin Fatih Vişneci MD, University of Health Sciences Türkiye, Konya City Hospital, Clinic of Emergency Medicine, Konya, Türkiye
E-mail: drfatihvisneci@hotmail.com **ORCID ID:** orcid.org/0000-0001-8050-4433
Cite this article as: Vişneci EF. Clinical profile and prognostic indicators of papaver rhoeas poisoning in adults: a retrospective analysis from Türkiye. Eurasian J Emerg Med. 2025;24(3): 228-31.

Received: 21.07.2025
Accepted: 21.08.2025
Published: 10.09.2025

Figure 1. Poppy Grass and- Poppy flower

on corn poppy toxicity primarily includes pediatric case reports (8).

However, its potential to cause severe toxic effects in adults has also been noted. Lactate levels have been reported as prognostic indicators in poisoning cases, although studies specifically addressing this in Papaver rhoeas poisoning are scarce (9,10). This study aims to contribute to the literature by analyzing the symptom profiles, time of presentation, lactate levels, and length of hospital stay in adult patients with corn poppy poisoning and examining their relationship with clinical severity.

Materials and Methods

The retrospective study was intended to be conducted at the emergency department of a tertiary care training and research hospital between January 2024 and May 2025. A total of 15 patients who were diagnosed with Papaver rhoeas (corn poppy) poisoning were included. All patients had consumed the poppy flower either as herbal tea for health or as a food by adding the flower's leaves into their meals before the onset of symptoms. Clinical records and laboratory results were obtained from the hospital's digital archive. Patients were categorized into two groups based on the severity of symptoms: mildly symptomatic (n=8) and severely symptomatic (n=7). The mild group included cases presenting with gastrointestinal symptoms or mild neurological manifestations, while the severe group consisted of patients exhibiting serious neurological symptoms such as seizures, convulsions, or loss of consciousness. The following parameters were evaluated: gastrointestinal, neurological, and cardiac symptoms; latency between ingestion and symptom onset (<1 hour vs. >1 hour); serum lactate level (<2 mmol/L vs. >2 mmol/L); and length of hospital stay (<48 hours vs. >48 hours).

The study was approved by the local ethics committee of University of Health Sciences Türkiye Konya City Hospital (desicion number: 11-67, date: 12.06.2025).

Statistical Analysis

The data analysis was performed using the IBM SPSS Statistics 21.0 (IBM Corp, Armonk, N.Y., USA) software package. Descriptive statistics were used, and Fisher's exact test was performed for categorical variables due to the small sample size. The Mann-Whitney U test was applied for comparisons of symptom scores between groups. A p value of <0.05 was considered statistically significant. Group comparisons were visualized graphically.

Results

The median age of the participants in the study was 60 (Q1=44, Q3=69); with 66.67% (n=10) being female and 33.33% (n=5) being male.

Among the 15 patients included in the study, 8 were classified as mildly symptomatic and 7 as severely symptomatic. The findings suggest that the toxic effects of corn poppy primarily involve the central nervous and gastrointestinal systems in severe cases, with limited cardiac involvement (Figure 2). Patients in both groups had no predisposing diseases in their history. Symptom distribution, time to presentation, lactate levels, and hospitalization durations are summarized below (Table 1). In the severe group, gastrointestinal symptoms, as well as mild and

severe neurological symptoms, were frequently observed, while no cardiac symptoms were documented. Time to presentation (p=0.041), serum lactate levels (p=0.033), and length of hospital stay (p=0.037) demonstrated statistically significant differences between the patient groups. Patients in this group, who were classified as severe, presented later (>1 hour after ingestion), having elevated lactate levels (>2 mmol/L), and experiencing longer hospitalizations (>48 hours) (Figure 3).

In contrast, symptom scores exhibited a markedly stronger distinction between the groups (p=0.002). Patients in this group were classified as mild, were presented earlier (<1 hour), had lower lactate levels (<2 mmol/L), and required shorter hospitalization (<48 hours). The presence of severe neurological symptoms and high lactate levels was strongly associated with poor prognosis (Figure 4).

Discussion

This study represents one of the few clinical analyses evaluating symptom severity, time to presentation, lactate levels, and hospitalization duration in patients admitted to the emergency department with Papaver rhoeas (corn poppy) poisoning. The findings suggest that, contrary to its common perception as a harmless herbal remedy, corn poppy may lead to significant toxicity (11). Most patients in the severe group presented with prominent neurological symptoms, which can be attributed to the isoquinoline alkaloids found in Papaver species that affect the central nervous system. The high frequency of gastrointestinal symptoms in both mild and severe groups indicates systemic toxicity. Interestingly, cardiac symptoms were observed only in a few mild cases and were absent in the severe group (12-15). Late presentation (>1 hour post-ingestion) and elevated serum lactate levels (>2 mmol/L) were significantly associated with severe poisoning. This aligns with previous literature emphasizing the prognostic value of lactate in various toxicological conditions. In our study, high lactate levels were correlated with longer hospital stays and more severe neurological manifestations, highlighting their clinical utility for early risk stratification in corn poppy toxicity (16,17). While most existing reports on corn poppy poisoning are isolated case studies, this study provides a comparative clinical analysis of adult patients. Another notable observation was that the majority of cases occurred during early spring (late March to early May), corresponding with the early sprouting phase of the plant before flowering. Traditional beliefs suggest that corn poppy should be consumed before blooming, which may have a biochemical basis: toxic alkaloid concentrations could vary depending on the plant's developmental stage (18). This hypothesis is supported by analogies to mushroom toxicity, in which biochemical

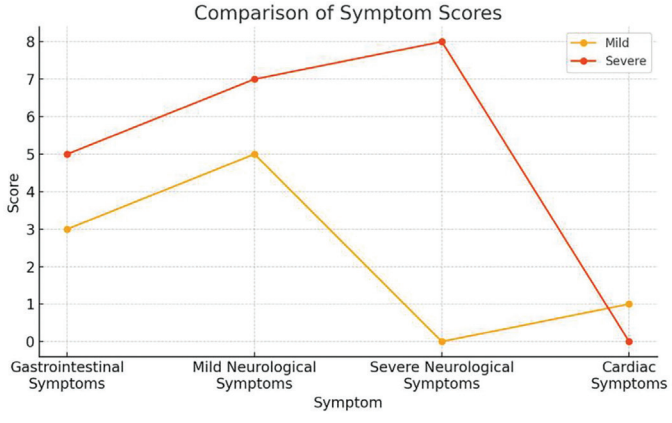


Figure 2. Comparison of symptom scores

Table 1. Comparative analysis of patient groups	
Time to presentation (<1 h vs >1 h)	p value=0.041
Lactate level (<2 vs >2)	p value=0.033
Hospital stay (<48h vs >48h)	p value=0.037
Total symptom score (mild vs severe)	p value=0.002

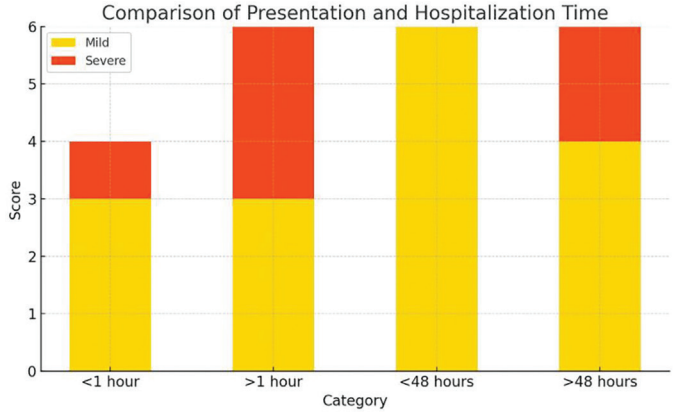


Figure 3. Comparison of presentation and hospitalization time

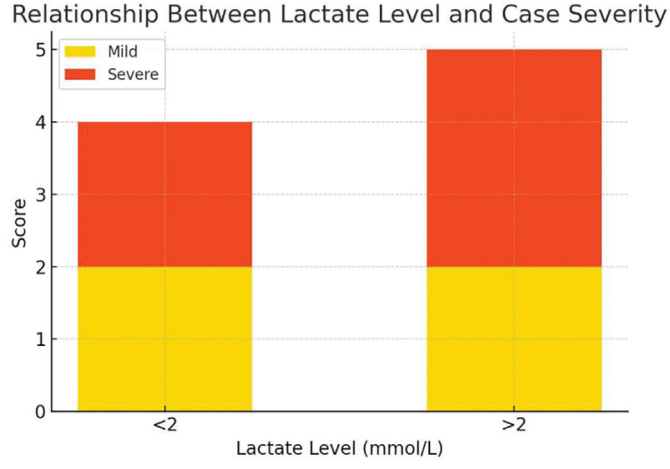


Figure 4. Relation between lactate level and case severity

composition changes with growth phases. Future prospective studies should investigate the phytochemical profile of Papaver rhoeas across different growth stages (e.g., sprouting, flowering, and seed phases) to identify the specific compounds responsible for toxicity. Chromatographic and molecular analyses performed in university-based botany and pharmacognosy laboratories could provide valuable contributions to the scientific literature.

Study Limitations

The limited number of cases in our study can be attributed to the seasonal variability of the poppy flower, which affects both its availability and the clinical presentation of related toxicological cases. Although some variables reached statistical significance, the small sample size limits the generalizability and power of the results. There is no specific test to determine the severity of Papaver rhoeas poisoning.

Conclusion

Although rare, corn poppy poisoning can lead to serious clinical manifestations in emergency settings. Delayed presentation, elevated lactate levels, and neurological symptoms are key indicators of poor prognosis. The public perception that herbal products are inherently safe must be re-evaluated, and healthcare professionals should be educated about the potential toxicity of traditional medicinal plants. This study also suggests that toxicity severity may not solely depend on the quantity ingested but also on the plant's developmental stage. Milder symptoms were typically observed with early spring consumption (pre-bloom), while more severe outcomes were associated with ingestion closer to the flowering stage. Future research should focus on chemical analyses of the plant at various growth stages to determine which phases present increased toxicological risks.

Ethics

Ethics Committee Approval: The study was approved by the local ethics committee of University of Health Sciences Türkiye Konya City Hospital (desicion number: 11-67, date: 12.06.2025).

Informed Consent: The retrospective study was intended to be conducted at the emergency department of a tertiary care training and research hospital between January 2024 and May 2025.

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. Gürbüz İ, Gençler Ö, Akaydın G, Salihoğlu E, Günbatan T, Demirci F, et al. Folk medicine in Düzce Province (Turkey). Turkish Journal of Botany. 2009;43:769-84.
2. Sarı AO, Oğuz B, Bilgiç A, Tort N, Güvensen A, Şenol, SG. Ege ve Güney Marmara bölgelerinde halk ilacı olarak kullanılan bitkiler. Anadolu Ege Tarımsal Araştırma Enstitüsü Dergisi. 2010;20:1-21.
3. Ceylan F, Akar Şahingöz S. Yenilebilir otlar tüketim alışkanlığı: düziçi örneği. Journal of Tourism and Gastronomy Studies. 2019;7:2204-25.
4. Raman SV, Jacob J. Mydriasis due to datura inoxia. Emerg Med J. 2005;22:310-1.
5. Balabanli C, Albayrak S, Türk M, Yüksel O, Some weeds are found in Turkey's pasture and its impact on animals, Journal of the Faculty of Forestry. 2006;2:89-96.
6. Koçak S, Karabulut K, Ertekin B, Nak H, Cander B. Red poppy (Papaver rhoeas) poisoning: a report of three cases. Cyprus J Med Sci. 2016;1:11-3.
7. Zargari A. Medicinal plants. Tehran University of Medical Sciences. 1997.
8. Amathieu R, Levesque E, Merle JC, Chemit M, Costentin C, Compagnon P, et al. Severe toxic acute liver failure: etiology and treatment. Ann Fr Anesth Reanim. 2013;32:416-21.
9. Deng JF. Clinical and laboratory investigations in herbal poisonings. Toxicology. 2002;181-182:571-6.
10. Soulimani R, Younos C, Jarmouni-Idrissi S, Bousta D, Khallouki F, Laila A. Behavioral and pharmaco-toxicologicalstudy of Papaverrhoeas L. in mice. J Ethnopharmacol 2001;74:265-74.
11. Kültür S. Medicinal plants used in Kirklareli province (Turkey). J Ethnopharmacol 2007;111:341-4.
12. Adukauskiene D, Dockiene I, Naginiene R, Kevelaitis E, Pundzius J, Kupcinskas L. Acuteliver failure in lithuania. Medicina (Kaunas). 2008;44:536-40.
13. Jóhannsson M, Ormarsdóttir S, Olafsson S. Hepatotoxicity associated with the use of Herbalife. Laeknabladid. 2010;96:167-72.
14. Khashab M, Tector AJ, Kwo PY. Epidemiology of acute liver failure. Curr Gastroenterol Rep. 2007;9:66-73.
15. Stickel F, Kessebohm K, Weimann R, Seitz HK. Review of liver injury associated with dietary supplements. Liver Int. 2011;31:595-605.
16. Kalav YN, Sariyar G. Alkaloids from Turkish Papaver rhoeas. Planta Med. 1989;55:488.
17. Pourmotabbed A, Rostamian B, Manouchehri G, Pirzadeh-Jahromi G, Sahraei H, Ghoshooni H, et al. Effects of Papaver rhoeas extract on the expression and development of morphine-dependence in mice. J Ethnopharmacol. 2004;95:431-5.
18. A poster showing three different stages of a Common Poppy flower (Papaver rhoeas): bud, flower and fruit (capsule). Avaible from: https://tr.wikipedia.org/wiki/Dosya:Poster_papaver_3a.jpg

EURASIAN JOURNAL OF EMERGENCY MEDICINE
Letter to the Editor
DOI: 10.4274/eajem.galenos.2025.48264
Eurasian J Emerg Med. 2025;24(3): 232-3

Prehospital Electrocardiography Training: Cross-Industry Standard Process for Data Mining Method or Traditional Method?

Göknur Yıldız

University of Health Sciences Türkiye, Eskişehir City Hospital, Clinic of Emergency Medicine, Eskişehir, Türkiye

Keywords: ECG, prehospital care, cross-industry standard process for data mining

Dear Editor,

I read the article titled effect of cross-industry standard process for data minin (CRISP) method training on electrocardiography (ECG) diagnosis skills of prehospital medical services personnel by Ekici et al. (1) with great interest, and I would like to address some points that merit more attention. ECG assessment is an important aspect in prehospital care. Unfortunately, many healthcare personnel have limited knowledge of ECG, and they should receive regular training in order to detect fatal and emergency-related ECG changes (2).

In their study, Ekici et al. (1) evaluated the effectiveness of the CRISP method and divided the study population into control and experimental groups. The analysis showed that the CRISP method was more successful in recognizing sinus rhythm, supraventricular tachycardia, atrial fibrillation, second-degree Mobitz type-1, and Mobitz type-2 atrioventricular blocks. However, no statistically significant difference was found between the two groups for sinus bradycardia, second-degree Mobitz type-2 atrioventricular block, atrial flutter, third-degree atrioventricular block, ventricular tachycardia, ventricular fibrillation, and asystole. Studies have shown that third-degree atrioventricular block, pulseless ventricular tachycardia, and ventricular fibrillation are among the leading causes of sudden cardiac death, and that a quick diagnosis and initiation of treatment are crucial (3,4). The fact that the CRISP method did not show superiority in these conditions can be considered a limitation of the study in terms of evaluating prehospital emergency cases.

Moreover, the ischemic findings such as ST-segment elevation and depression, which should be quickly identified in prehospital ECG, were not evaluated in the study by Ekici et al. (1). As mentioned in the introduction of the study, early detection of acute cardiac ischemic conditions in the prehospital setting is important for ensuring appropriate hospital transfer and facilitating rapid cardiac catheterization (5). If the study had included evaluations of ST-segment changes in the ECG, it could have contributed more meaningfully to the literature and yielded more appropriate results.

Using traditional or contemporary methods, providing ECG training to healthcare personnel working in prehospital medical services at regular intervals will automatically improve their ECG assessment capabilities.

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.

Corresponding Author: Göknur Yıldız MD, University of Health Sciences Türkiye, Eskişehir City Hospital, Clinic of Emergency Medicine, Eskişehir, Türkiye
E-mail: goknur_yldz@hotmail.com **ORCID ID:** orcid.org/0000-0002-2097-2531

Cite this article as: Yıldız G. Prehospital electrocardiography training: cross-industry standard process for data minin method or traditional method? Eurasian J Emerg Med. 2025;24(3): 232-3.

Received: 06.01.2025
Accepted: 07.02.2025
Epub: 20.05.2025
Published: 10.09.2025

References

1. Ekici M, Turan ÖF, Borazan İ, Aksoy FM, Yaman S, Kaplan Z. Effect of CRISP method training on ECG diagnosis skills of prehospital medical services personnel. Eurasian J Emerg Med. 2024;23:257-62.
2. Breen CJ, Kelly GP, Kernohan WG. ECG interpretation skill acquisition: a review of learning, teaching and assessment. J Electrocardiol. 2022;73:125-8.
3. Rea TD, Pearce RM, Raghunathan TE, Lemaitre RN, Sotoodehnia N, Jouven X, et al. Incidence of out-of-hospital cardiac arrest. Am J Cardiol. 2004;93:1455-60.
4. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, et al. Heart disease and stroke statistics-2018 update: a report from the american heart association. Circulation. 2018;137:e67-e492.
5. Mitsis A, Gragnano F. Myocardial Infarction with and without ST-segment elevation: a contemporary reappraisal of similarities and differences. Curr Cardiol Rev. 2021;17:e230421189013.

72 TH EPAT CARDIAC LIFE SUPPORT (CLS)

27-28 SEPTEMBER 2025
BEZMIALEM VAKIF UNIVERSITY / ISTANBUL / TÜRKİYE

Director
Prof. Dr. Başar CANDER

Coordinator
Prof. Dr. Zeynep ÇAKIR

72 th Epat Cardiac Life Support Course will be held in Istanbul on 27-28 Sep 2025.

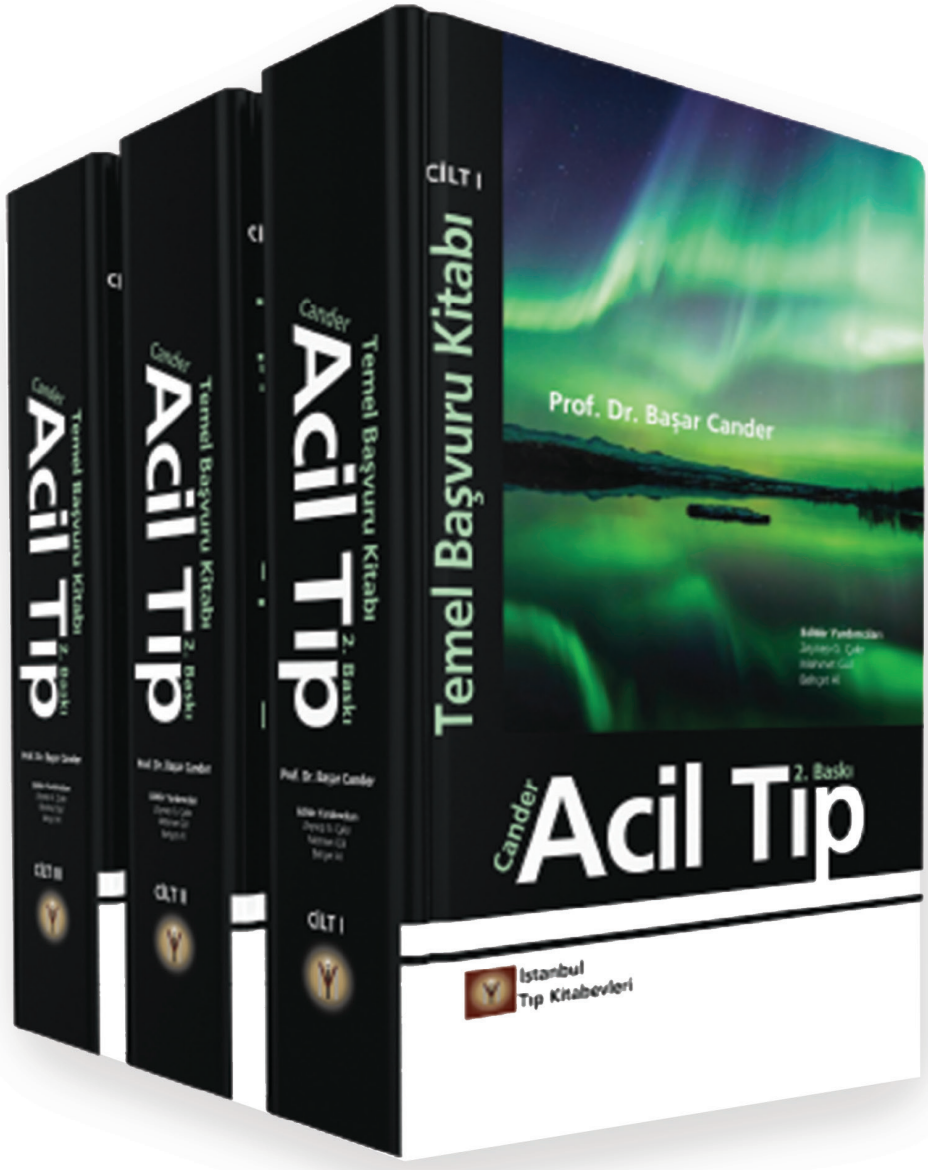
You can follow the course registration conditions at www.atuder.org.tr.

A registration fee must be paid for this course. Attendance is mandatory for this course, which will last 2 full days and will include theoretical training until noon each day, and one-on-one practical training with devices on models in the afternoon, and participants who pass the exam at the end of the course will be given a "Certificate of Achievement" valid for 3 years.

EPAT
Emergency Physicians Association of Turkey
Adres: Güleryüz Sok. Emek Apt. No:26/19 Y. Ayrancı / Çankaya / Ankara
T: 0312 426 12 14 • F: 0312 426 12 44
@atuder1999 @atuderorg @atuder_epat @atuder1999 info@atuder.org.tr atuder.org.tr

EMERGENCY MEDICINE IS OUT NOW!

EMERGENCY MEDICINE BASIC CONSULTING BOOK WITH THE SUPPORT OF İSTANBUL TIP KİTABEVLERİ



www.istanbultip.com.tr

Critical Care

Editors: **Başar Cander, Prof. MD | Lukasz Szarpak, Prof. MD**
Co-Editors: **Mehmet Gül, Prof. MD | Hassan Al-Thani, MD**
Ayman El-Menyar, MD

İstanbul
Tıp Kitabevi