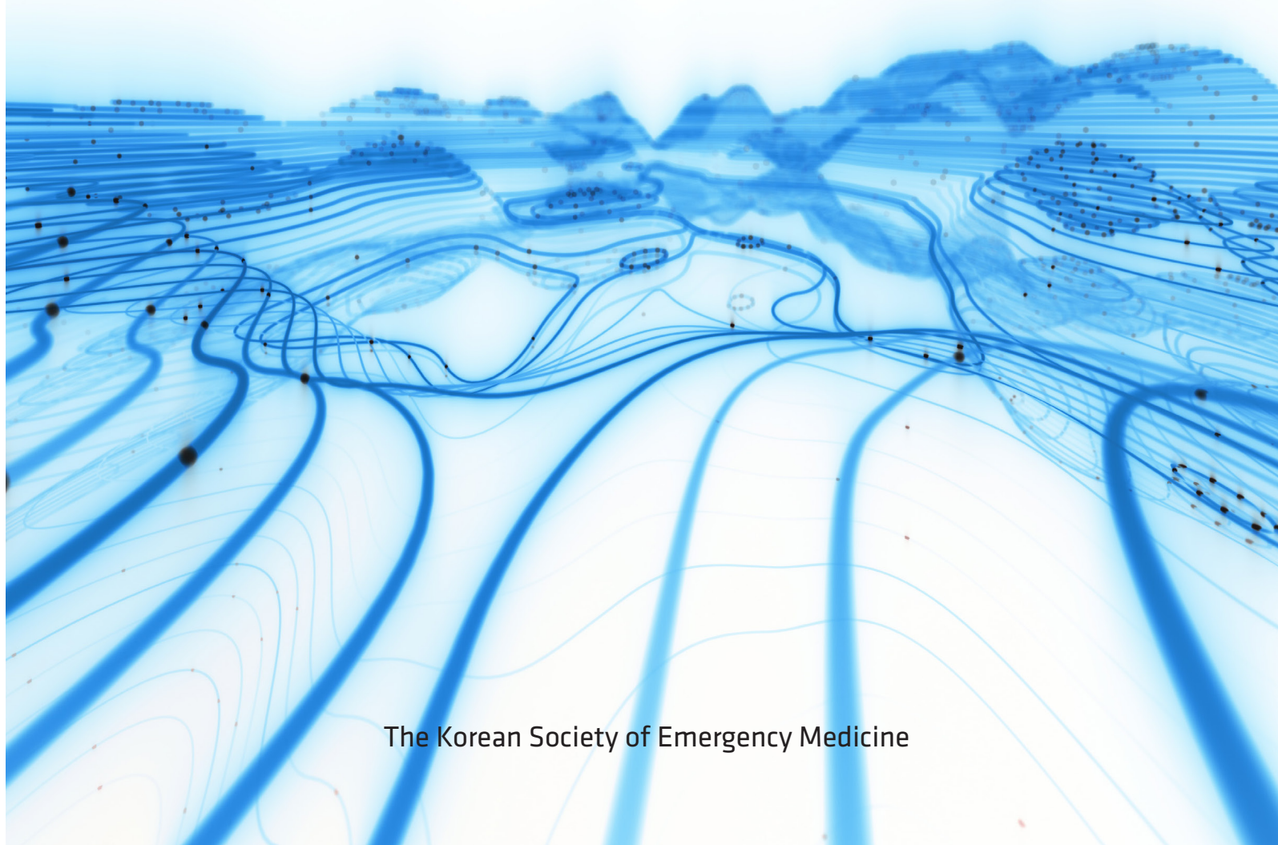
Clinical and Experimental Emergency Medicine

pISSN: 3022-1811
eISSN: 2383-4625

Clinical and Experimental Emergency Medicine

Vol. 13 • No. 1 • March 2026

www.ceemjournal.org



The Korean Society of Emergency Medicine

CEEM Clinical and Experimental Emergency Medicine

Vol. 13 • No. 1 • March 2026

pages 1-111

Clinical and Experimental Emergency Medicine

www.ceemjournal.org

Vol. 13 • No. 1 • March 2026

pISSN: 3022-1811
eISSN: 2383-4625

Published on March 31, 2026

Clinical and Experimental Emergency Medicine (CEEM) is the official journal of the Korean Society of Emergency Medicine. CEEM was launched on September 30, 2014, and is published quarterly on the last day of March, June, September, and December, one volume per year. As a peer-reviewed, open access journal, CEEM focuses on both basic and clinical research of emergency medicine including pathophysiology, epidemiology, diagnosis, prognosis, treatment, and simulation. CEEM publishes editorials, original articles, reviews, study protocol, letters to the editor, case reports, images in emergency medicine, and more. CEEM will be of interest to healthcare professionals in acute care and emergency medicine, pediatric emergency medicine, emergency medical services, emergency procedures, cardiology, neurology, resuscitation, trauma, education, and so on. CEEM is one of the only journals that covers basic and clinical research fields entirely focusing on acute care and emergency medicine. All or part of the journal contents are indexed or covered by PubMed, PubMed Central, Scopus, Embase, Emerging Sources Citation Index (ESCI), Korea Citation Index (KCI), Google Scholar, and CrossRef. Full text is also freely available from the journal website (https://www.ceemjournal.org/). CEEM was supported by the Korean Federation of Science and Technology Societies (KOFST) grant, funded by the Korean government.

Open Access

CEEM is an open access journal distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Instructions to Authors are available from the journal website (https://www.ceemjournal.org/authors/authors.php).

Publisher

The Korean Society of Emergency Medicine

Editor-in-Chief

Adam J. Singer, Stony Brook University, USA
Kyuseok Kim, CHA University, Korea

Editorial Office

101-3104, Brownstone Seoul, 464 Cheongpa-ro, Jung-gu, Seoul 04510, Korea
Tel: +82-31-709-0918, Email: office@ceemjournal.org

Printing Office

M2PI
805, 26 Sangwon 1-gil, Seongdong-gu, Seoul 04779, Korea
Tel: +82-2-6966-4930, Fax: +82-2-6966-4945, Email: support@m2-pi.com

♾ This paper meets the requirements of KS X ISO 9706, ISO 9706-1994, and ANSI/NISO Z39. 48-1992 (permanence of paper).

©The Korean Society of Emergency Medicine

Editorial Board

Vol. 13 • No. 1 • March 2026

Clinical and Experimental Emergency Medicine

pISSN: 3022-1811 | eISSN: 2383-4625

Editor-in-Chief

Adam J. Singer
Stony Brook University, USA

Kyuseok Kim
CHA University, Korea

Senior Editor

Clifton Callaway	University of Pittsburgh, USA
David Cone	Yale University, USA
Colin A. Graham	The Chinese University of Hong Kong, China
Sung Oh Hwang	Yonsei University Wonju College of Medicine, Korea
Judd Hollander	Thomas Jefferson University, USA
Nathan Kuppermann	University of California, Davis, USA
Marcus Ong Eng Hock	National University of Singapore, Singapore
Kyu Nam Park	The Catholic University of Korea, Korea
W. Frank Peacock	Baylor College of Medicine, USA
Nathan Shapiro	Harvard University, USA
Ian G. Stiell	University of Ottawa, USA
Gil Joon Suh	Seoul National University, Korea
Wanchun Tang	Weil Institute of Critical Care Medicine, USA
Jean-Louis Vincent	Université Libre de Bruxelles, Belgium

Deputy Editor

Tae Gun Shin	Sungkyunkwan University, Korea

Associate Editor

Bentley Bobrow	University of Arizona, USA
Wonchul Cha	Sungkyunkwan University, Korea
Sangchun Choi	Soonchunhyang University, Korea
Sung-Hyuk Choi	Korea University, Korea
Sung Phil Chung	Yonsei University, Korea
Deborah Diercks	University of California, Davis, USA
Michael Gottlieb	Rush Medical Center, USA
Pinchas Halpern	Tel-Aviv University, Israel
Andy Jagoda	Mount Sinai, USA
You Hwan Jo	Seoul National University, Korea
Alan E. Jones	University of Mississippi, USA
Won Young Kim	University of Ulsan, Korea
Andrew Kirkpatrick	University of Calgary, Canada
Patrick Ko	National Taiwan University, Taiwan
Jae Woo Lee	University of California, San Francisco, USA
Ki Won Lee	The University of Texas, USA
Jisook Lee	Ajou University, Korea
Kang Hyun Lee	Yonsei University Wonju College of Medicine, Korea
Tae Ho Lim	Hanyang University, Korea
Christopher Lipinski	Mayo Medical School, USA
James Miner	University of Minnesota, USA
Vicki Noble	Harvard University , USA
Sang Do Shin	Seoul National University, Korea
Kjetil Sunde	University of Oslo, Norway
Scott D. Weingart	Stony Brook University, USA
Young Hoon Yoon	Korea University, Korea

©The Korean Society of Emergency Medicine

Editorial Board

Vol. 13 • No. 1 • March 2026

Clinical and Experimental Emergency Medicine

pISSN: 3022-1811 | eISSN: 2383-4625

Section Editor

Airway

Sung Yeon Hwang — Sungkyunkwan University, Korea

Critical Care

You Hwan Jo — Seoul National University, Korea

Byuk Sung Ko — Hanyang University, Korea

Digital healthcare & informatics

Taerim Kim — Sungkyunkwan University, Korea

Education & Simulation

Jaehoon Oh — Hanyang University, Korea

EMS & Disaster

Jeong Ho Park — Seoul National University, Korea

Experimental research

Jae Hyuk Lee — Seoul National University, Korea

Imaging

Young Soon Cho — Soonchunhyang University, Korea

Changsun Kim — Hanyang University, Korea

Medical Emergency

Shin Ahn — University of Ulsan, Korea

Tae Nyoung Chung — CHA University, Korea

Pediatrics

Jisook Lee — Ajou University, Korea

Public health & Injury prevention

Young Sun Ro — Seoul National University, Korea

Procedure

Jung Youn Kim — Korea University, Korea

Resuscitation

Byung Kook Lee — Chonnam National University, Korea

Je Hyeok Oh — Chung-Ang University, Korea

Jung Soo Park — Chungnam National University, Korea

Toxicology

Sangchun Choi — Soonchunhyang University, Korea

Yong Sung Cha — Yonsei University Wonju College of Medicine

Trauma

Young Hoon Yoon — Korea University, Korea

Editorial Board Member

Jin Ho Beom — Yonsei University, Korea

Jun Hwi Cho — Kangwon National University, Korea

Jin Woo Jeong — Dong-A University, Korea

Kyung Woon Jeung — Chonnam National University, Korea

Soo Kang — Inha University, Korea

Taeyoung Kong — Yonsei University, Korea

Woon Yong Kwon — Seoul National University, Korea

Sung Woo Lee — Korea University, Korea

Mi Jin Lee — Kyungpook National University, Korea

Sang Hoon Oh — The Catholic University of Korea, Korea

Joo Suk Oh — The Catholic University of Korea, Korea

Sang O Park — Konkuk University, Korea

Jae-Hyug Woo — Gachon University, Korea

Jae Chol Yoon — Jeonbuk National University, Korea

Je Sung You — Yonsei University, Korea

Chun Song Youn — The Catholic University of Korea, Korea

Statistical Editor

Seung Sik Hwang — Seoul National University,Korea

Manuscript Editor

Seeun Hur — InfoLumi, Korea

Editorial Assistant

Hayoung Kim — InfoLumi, Korea

©The Korean Society of Emergency Medicine

CEEM
Clinical and Experimental Emergency Medicine

Vol. 13 • No. 1 • March 2026

pISSN: 3022-1811
eISSN: 2383-4625

Editorial

1 **Artificial intelligence in emergency medicine critical care**
Samita M. Heslin, Robert Nocito, Scott D. Weingart

Review Articles

5 **Initial diagnosis and management of acute ischemic stroke: updates and future directions**
Sang Hee Ha, Sujin Koo, Seung Min Kim, Jae Young Park, Seung Eun Lee, Bum Joon Kim

13 **The role of fomepizole in acetaminophen-related poisoning: a narrative review**
Frank Chen, Shelly ZQ Lu, Harleen Chohan, Anthony Lau

28 **Public awareness of medical emergency telephone numbers: a scoping review**
Alexei A. Birkun

Original Articles

37 **Epidemiologic trends in pediatric urinary tract infections in United States emergency departments from 2016 to 2023**
Michael Gottlieb, Tamara Amponsah, Nhat Nguyen, Ohm Shukla, Kyle Bernard, Eric Moyer

44 **Comparison of four noninvasive tools for predicting sepsis and septic shock mortality: a prospective cohort study**
Matteo Guarino, Giacomo Maroncelli, Benedetta Perna, Paolo Baldin, Caterina Ghirardi, Alex Zanotto, Silvia Eichner, Matteo Bolognesi, Chiara Pesci, Martina Maritati, Carlo Contini, Roberto De Giorgio, Michele Domenico Spampinato

53 **Automated chain-of-thought evaluation framework for large language model–generated emergency department documentation: a simulation-based study**
Dasol Choi, Junhyuk Seo, Won Cul Cha, Minha Kim, Sejin Heo, Hansol Chang, Taerim Kim

65 **Skull fractures may be associated with hyperfibrinolysis in patients with isolated traumatic brain injury**
Gaku Fujiwara, Naoto Shiomi

74 **A pilot study of recurrent abdominal pain in the emergency department: low-risk disease associated with high pain severity and frequent opioid administration**
Soroush Shahamatdar, Felice Yang, Meylakh Barshay, Ryan Heidish, Aditya Loganathan, Taylor Bolden, William Huang, Andrew C. Meltzer

Vol. 13 • No. 1 • March 2026

pISSN: 3022-1811
eISSN: 2383-4625

Brief Research Report

81 **Performance of a transesophageal echocardiography probe at temperature monitoring during simulated hypothermia and rewarming**
Madeline Larochelle, Margarita Popova, David Mackenzie, Andrew Fried, Peter Croft, Joshua Rehberg, Christina Wilson

Case Report

86 **Mortality due to minimal bentazone intoxication: a case report**
Yumin Jeon, Sejoong Ahn, Jong-Hak Park, Hanjin Cho, Sungwoo Moon, Sukyo Lee

Commentary

90 **Epidemiology of cardiac arrest and acute myocardial infarction in emergency departments in Korea, 2020–2024: a report using the National Emergency Department Information System (NEDIS) database**
Sejoong Ahn, Jung Eon Kim, Sung-keun Ko, Eunsil Ko

98 **Emergency department presentations for suicide and self-harm in Korea, 2020–2024: an epidemiological study using the National Emergency Department Information System (NEDIS) database**
Yuri Choi, Jinwoo Jeong, Borami Lim, Myeong Il Cha

105 **Emergency medicine physicians treating urine cultures when it is not clear they should**
Amanda Doodlesack, Clifford Marks, Nicolette Castaneda, Jonathan Edlow, Shamai A. Grossman

Images in Emergency Medicine

110 **A false-positive Focused Assessment with Sonography in Trauma (FAST) in a trauma patient**
Lindsay Taylor, Fred Milgrim, Matthew Riscinti, Michael Gottlieb

© 2026 The Korean Society of Emergency Medicine

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):1-4
https://doi.org/10.15441/ceem.26.057

Artificial intelligence in emergency medicine critical care

Samita M. Heslin[1], Robert Nocito[1], Scott D. Weingart[2]

[1]Department of Emergency Medicine, Stony Brook University, Stony Brook, NY, USA
[2]Department of Emergency Critical Care, Nassau University Medical Center, East Meadow, NY, USA

pISSN: 3022-1811
eISSN: 2383-4625

Editorial

INTRODUCTION

In emergency medicine (EM) critical care, patients may arrive clinically unstable, with limited information located in multiple medical records, and require critical interventions before their clinical course becomes clear. Decision-making is time-sensitive and patients' conditions may continually change in response to treatment.

Interest in artificial intelligence (AI) in critical care has grown with advances in machine learning, the availability of longitudinal data, and the encouraging performance of predictive models in retrospective intensive care unit (ICU) studies [1,2]. Much of this work has been conducted in the traditional ICU settings and has provided foundational insights. However, emergency critical care presents important challenges for many existing AI methods and needs systems to be evaluated and adapted for this environment. Understanding why AI implementation in EM critical care differs from that in other areas of medicine is important for effective AI integration in this setting.

CHALLENGES

AI models for early deterioration prediction often use continuous physiologic and laboratory data to predict clinical decline [2,3]. However, structured modelling of dynamic changes from acute interventions such as airway management, vasopressors initiation, fluid resuscitation, or procedural treatments, which occur frequently in emergency critical care, remains relatively limited in many current approaches.

Early predictions may need to change as the patient's condition continues to change over time [4,5]. A mortality or deterioration risk assessment performed prior to airway management may not accurately predict the patient's physiological status after intubation. Similarly, hemodynamic predictions made prior to initiating vasoactive support may not accurately represent the patient's cardiovascular physiology after treatment.

Therefore, AI tools designed for the EM critical care environment should incorporate time awareness and continuously update outputs as the patient's physiology and treatment continue to change. Instead of providing a single forecast, AI is more beneficial when tracking physiological trends over time, identifying changes in the patient's clinical status, and signaling to the clinician when previous assessments may no longer apply [4,5].

Emergency critical care patients are often given working diagnoses early in their course to guide initial management. However, these diagnoses are frequently revised as the patient responds to treatment and when additional data and clinical patterns develop.

Received: February 10, 2026
Revised: February 27, 2026
Accepted: February 28, 2026

Correspondence to:
Samita M. Heslin
(SamitaHeslin@gmail.com)

How to cite this article:
Heslin SM, Nocito R, Weingart SD. Artificial intelligence in emergency medicine critical care. Clin Exp Emerg Med 2026;13(1):1-4. https://doi.org/10.15441/ceem.26.057

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

© 2026 The Korean Society of Emergency Medicine

AI tools in the EM critical care environment should focus less on classifying diagnoses and more on recognizing and describing evolving patient states. This could aid clinicians in assessing how the patient's physiology is responding to therapy and how new risks are developing. Systems that can analyze and display multiple vital signs, laboratory trends, and ventilatory parameters over time may be better suited for caring for critically ill patients.

Uncertainty is also part of emergency critical care. Clinicians must develop plans of action with incomplete data and weigh the risk of potential complications against the risk of the patient's condition deteriorating. Some AI systems may report results that appear confident despite underlying uncertainty [6]. When AI system outputs appear to be authoritative, clinicians may be reluctant to challenge them [7,8].

APPLICATIONS OF AI IN EM CRITICAL CARE

Despite the limitations, AI systems may hold value in the emergency critical care environment when aligned with clinical workflows and the dynamic physiology of the patient.

Continuous longitudinal physiologic monitoring

AI can help clinicians with analyzing information across multiple notes, prior records, and documented goals of care. It could also help monitor physiological trends over time rather than relying on isolated data points. Machine learning models are well-suited for identifying patterns in patients' continuous vital signs and combining data from multiple monitoring sources to identify early warning signs of patient decompensation [1,3,9–12]. In some cases, these systems generate alerts based on patterns of physiologic change that may not be readily apparent to clinicians, yet still accurately predict clinical decline [9,10]. The effective implementation of AI could help refocus monitoring environments to prioritize clinically significant alarms, reduce the frequency of false alarms, and provide earlier recognition of potentially life-threatening instability.

Treatment responsive decision support

Instead of producing predictions independent of care, AI systems should be designed to consider the interventions already initiated. Accounting for treatment context (e.g., recent intubation or fluid administration) can increase the applicability of AI-generated output and decrease the likelihood of misleading output [4,5,13].

Transition of care

Transition of care handoffs are important in emergency critical care and a recognized area of communication risk [14–16]. AI-generated summaries that show active issues, physiological changes, code status, and pending tasks can help with continuity of care. The key to successfully integrating AI-generated summaries into clinical workflows is to streamline the information, making it concise and focused on what is needed to support immediate clinical decisions.

INTEGRATING INTO THE ENVIRONMENT

For AI systems to be successfully integrated into the EM critical care environment, workflow integration, clinician trust, and education are important. AI tools that add extra steps to the workflow, require separate interfaces, or deliver excessive information will increase the cognitive burden on the clinician and defeat the purpose of implementing AI to reduce workload [17,18]. Clinicians' trust in AI systems will depend on the transparency of the AI system, continuous monitoring of its performance, and its responsiveness to feedback as clinical practices develop [2,17,18]. Education on AI should be viewed as a clinical skill and include training in interpreting AI-generated output, understanding AI system limitations, and maintaining the ability to make independent judgments [7,18].

Data drift is the gradual difference between the data distributions a model was trained on and the data it is exposed to when applied in practice. In EM critical care settings, data drift may occur due to variations in patient populations, differing local clinical protocols, variability in documentation practices, and use of differing medical equipment among hospitals [19,20]. When models trained on data collected within a development environment begin to fail when they are applied at another clinical site or when clinical practices change over time, models that were effective in one environment may perform poorly in other environments [21,22]. Addressing data drift requires monitoring a model's performance after deployment continuously, defining triggers for reevaluation or retraining, and engaging with local clinical champions who can provide a perspective on performance issues.

FUTURE DIRECTIONS

AI in medicine is moving towards the development of foundation models and multimodal systems that are capable of processing heterogeneous data streams. This evolution is also applicable to EM critical care. The ability of multimodal AI systems to concurrently process electrocardiogram data, imaging studies, vital sign data, laboratory results, and clinical documentation provides the

capability for the construction of a more comprehensive and dynamic view of the critically ill patient [23,24]. Instead of relying upon a single data modality to identify patterns, multimodal AI systems can identify relationships between data modalities that are individually ambiguous, but collectively informative. The integration of multimodal AI systems into EM critical care departments will depend on the degree to which these systems have been validated in this specific clinical environment [25,26].

As AI systems continue to develop in complexity, so too does the need for transparency and explainability. High-stakes clinical environments require careful scrutiny of black-box models that offer recommendations without showing how the outputs were generated and why the clinician should act on them [27,28]. Explainable AI (XAI) methods, such as revealing which individual data elements contributed the most to the model's output, can help clinicians understand how the model reached its recommendations [27–30]. Prioritizing the development of XAI methodologies is crucial as AI tools are developed and improved for EM critical care departments.

There are ethical and legal considerations that may influence the future development of AI. One ethical concern is algorithmic fairness, which aims to prevent AI systems from reinforcing existing inequities in healthcare. When AI systems are trained on historical data, they may reflect disparities. As a result, they may produce systematically different recommendations across racial, socioeconomic, and demographic subgroups [31,32]. Ensuring that AI systems operate fairly across a diverse range of patient populations will require explicit consideration of equity during development, validation, and deployment [31,32]. Another issue is determining accountability when an AI system is involved in an adverse outcome. For example, who is responsible for harm caused by an AI system [33]? Other considerations include data privacy, obtaining informed consent for AI-assisted care, and the possibility of reidentification of deidentified training data. Therefore, multidisciplinary governance will be essential to the design and governance of AI systems.

CONCLUSION

The EM critical care setting has the potential to demonstrate both how AI can be used successfully in a medical context as well as the limits of AI in medicine. In order for AI systems to support the dynamic and intervention-rich environment of EM critical care, they must be able to adapt to changing conditions, aware of the time constraints involved in providing acute care, and compatible with the bedside workflow that is typical of the EM critical care setting. Without careful validation and governance, they risk adding to, rather than reducing, the cognitive burden of clinicians.

ARTICLE INFORMATION

Conflicts of interest
The authors have no conflicts of interest to declare.

Funding
The authors received no financial support for this study.

Data availability
Data sharing is not applicable as no new data were created or analyzed in this study.

REFERENCES

1. Pinsky MR, Bedoya A, Bihorac A, et al. Use of artificial intelligence in critical care: opportunities and obstacles. Crit Care 2024;28:113.
2. Oh W, Veshtaj M, Sawant A, et al. ORAKLE: Optimal Risk prediction for mAke30 in patients with sepsis associated AKI using deep LEarning. Crit Care 2025;29:212.
3. Hyland SL, Faltys M, Hüser M, et al. Early prediction of circulatory failure in the intensive care unit using machine learning. Nat Med 2020;26:364–73.
4. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med 2022;28:1773–84.
5. Tomašev N, Glorot X, Rae JW, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 2019;572:116–9.
6. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388:1233–9.
7. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17.
8. Khera R, Simon MA, Ross JS. Automation bias and assistive AI: risk of harm from AI-driven clinical decision support. JAMA 2023;330:2255–7.
9. Chang H, Park JE, Lee D, et al. Development and validation of a transformer model-based early warning score for real-time prediction of adverse outcomes in the emergency department. Sci Rep 2025;15:23021.
10. Salehinejad H, Meehan AM, Rahman PA, Core MA, Borah BJ,

Caraballo PJ. Novel machine learning model to improve performance of an early warning system in hospitalized patients: a retrospective multisite cross-validation study. EClinicalMedicine 2023;66:102312.
11. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med 2018;46:547–53.
12. Zhang Y, Xu W, Yang P, Zhang A. Machine learning for the prediction of sepsis-related death: a systematic review and meta-analysis. BMC Med Inform Decis Mak 2023;23:283.
13. Buell KG, Spicer AB, Casey JD, et al. Individualized treatment effects of oxygen targets in mechanically ventilated critically ill adults. JAMA 2024;331:1195–204.
14. Starmer AJ, Spector ND, Srivastava R. et al. Changes in medical errors after implementation of a handoff program. N Engl J Med 2014;371:1803–12.
15. Hartman V, Zhang X, Poddar R, et al. Developing and evaluating large language model-generated emergency medicine handoff notes. JAMA Netw Open 2024;7:e2448723.
16. Song JW, Park J, Kim JH, You SC. Large language model assistant for emergency department discharge documentation. JAMA Netw Open 2025;8:e2538427.
17. Lukac PJ, Turner W, Vangala S, et al. Ambient AI scribes in clinical practice: a randomized trial. NEJM AI 2025;2:10.1056/aioa2501000.
18. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020;20:310.
19. Finlayson SG, Subbaswamy A, Singh K, et al. The clinician and dataset shift in artificial intelligence. N Engl J Med 2021;385:283–6.
20. Wong A, Otles E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med 2021;181:1065–70.
21. Kore A, Abbasi Bavil E, Subasri V, et al. Empirical data drift detection experiments on real-world medical imaging data. Nat Commun 2024;15:1887.
22. Subasri V, Krishnan A, Kore A, et al. Detecting and remediating harmful data shifts for the responsible deployment of clinical AI models. JAMA Netw Open 2025;8:e2513685.
23. Cecconi M, Greco M, Shickel B, et al. Implementing artificial intelligence in critical care medicine: a consensus of 22. Crit Care 2025;29:290.
24. Azarfar G, Naimimohasses S, Rambhatla S, et al. Responsible adoption of multimodal artificial intelligence in health care: promises and challenges. Lancet Digit Health 2025;7:100917.
25. Schouten D, Nicoletti G, Dille B, et al. Navigating the landscape of multimodal AI in medicine: a scoping review on technical challenges and clinical applications. Med Image Anal 2025;105:103621.
26. Kotula CA, Martin J, Carey KA, et al. Comparison of multimodal deep learning approaches for predicting clinical deterioration in ward patients: observational cohort study. J Med Internet Res 2025;27:e75340.
27. Abgrall G, Holder AL, Chelly Dagdia Z, Zeitouni K, Monnet X. Should AI models be explainable to clinicians? Crit Care 2024;28:301.
28. Okada Y, Ning Y, Ong ME. Explainable artificial intelligence in emergency medicine: an overview. Clin Exp Emerg Med 2023;10:354–62.
29. Rosenbacke R, Melhus Å, McKee M, Stuckler D. How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care: systematic review. JMIR AI 2024;3:e53207.
30. Abbas Q, Jeong W, Lee SW. Explainable AI in clinical decision support systems: a meta-analysis of methods, applications, and usability challenges. Healthcare (Basel) 2025;13:2154.
31. Hasanzadeh F, Josephson CB, Waters G, Adedinsewo D, Azizi Z, White JA. Bias recognition and mitigation strategies in artificial intelligence healthcare applications. NPJ Digit Med 2025;8:154.
32. Amiot F, Potier B. Artificial intelligence (AI) and emergency medicine: balancing opportunities and challenges. JMIR Med Inform 2025;13:e70903.
33. Mello MM, Guha N. Understanding liability risk from using health care artificial intelligence tools. N Engl J Med 2024;390:271–8.

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):5-12
https://doi.org/10.15441/ceem.25.241

Review Article

Initial diagnosis and management of acute ischemic stroke: updates and future directions

Sang Hee Ha[1*], Sujin Koo[2*], Seung Min Kim[3], Jae Young Park[4], Seung Eun Lee[4], Bum Joon Kim[4]

[1]Department of Neurology, Gachon University Gil Medical Center, Incheon, Korea
[2]Department of Emergency Medical Technology, Doowon Technical University, Paju, Korea
[3]Department of Neurology, Veterans Health Service Medical Center, Seoul, Korea
[4]Department of Neurology, Asan Medical Center, Seoul, Korea

pISSN: 3022-1811
eISSN: 2383-4625

Received: October 7, 2025
Revised: October 29, 2025
Accepted: November 5, 2025

Correspondence to:
Bum Joon Kim
(medicj80@hanmail.net)

*Sang Hee Ha and Sujin Koo contributed equally to this study as co-first authors.

Outcomes in acute ischemic stroke (AIS) depend critically on rapid and accurate early diagnosis in the emergency department. Traditional prehospital tools and large vessel occlusion–focused scales facilitate triage but have limited capacity to distinguish ischemic from hemorrhagic stroke, a distinction essential for acute-phase treatment decisions. Recent advances include mobile stroke units equipped with computed tomography (CT), point-of-care laboratories, and telemedicine systems, as well as the emergence of biomarkers that enable field-based diagnosis and faster initiation of therapy. In-hospital imaging strategies incorporating CT, CT perfusion, and magnetic resonance imaging (MRI)-based tissue clocks have expanded eligibility for endovascular thrombectomy to include patients with large-core infarction or unclear-onset wake-up strokes. Prolonged cardiac monitoring and high-resolution vessel wall MRI have improved the detection of embolic sources and high-risk atherosclerotic plaques. Artificial intelligence now supports rapid imaging interpretation, workflow optimization, and treatment selection. Tenecteplase, a novel thrombolytic, provides a practical alternative to alteplase with comparable safety and efficacy, while post-thrombectomy management emphasizes individualized blood pressure control. In patients with minor stroke or high-risk transient ischemic attack, short-term dual antiplatelet therapy reduces early recurrence, and early initiation of lipid-lowering agents after AIS may stabilize vulnerable plaques and enhance vascular outcomes. Collectively, these innovations represent a shift toward integrated, time-sensitive, and precision-based AIS care spanning prehospital assessment, emergency department management, and post-reperfusion management.

Keywords Acute ischemic stroke; Artificial intelligence; Prehospital care; Rapid diagnosis; Acute thrombolysis; Stroke management

INTRODUCTION

Few conditions in emergency medicine rival acute ischemic stroke (AIS) in the urgency of diagnosis and treatment. In emergent large vessel occlusion (LVO), each minute without reperfusion results in irreversible neuronal loss and worsened long-term outcomes [1]. For emergency physicians, the first hours represent not only a race against time but also a critical window of opportunity in which rapid, well-informed decisions can dramatically alter a patient's outcome—from severe disability to functional independence.

Historically, prehospital and emergency department (ED) stroke assessment has relied on clini-

How to cite this article:
Ha SH, Koo S, Kim SM, Park JY, Lee SE, Kim BJ. Initial diagnosis and management of acute ischemic stroke: updates and future directions.
Clin Exp Emerg Med 2026;13(1):5-12.
https://doi.org/10.15441/ceem.25.241

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

© 2026 The Korean Society of Emergency Medicine

Capsule Summary

What is already known

Outcomes in acute ischemic stroke are highly time-dependent, with rapid and accurate diagnosis being essential for effective treatment. Conventional prehospital triage tools aid in identifying stroke symptoms and assessing severity but have limited ability to distinguish ischemic from hemorrhagic stroke or to guide treatment decisions beyond standard time windows. Intravenous alteplase has long been the standard thrombolytic therapy, while endovascular thrombectomy is firmly established for large vessel occlusion within conventional eligibility criteria. In addition, secondary prevention through antiplatelet therapy and lipid-lowering strategies has traditionally been emphasized from the acute phase through the chronic phase.

What is new in the current study

Recent advances are transforming acute stroke care toward a more integrated, data-driven, and precision-based model. Mobile stroke units, point-of-care biomarkers, and telemedicine now enable field-based diagnostic confirmation and earlier initiation of therapy. Imaging innovations, including computed tomography– and magnetic resonance imaging (MRI)-based tissue clocks, have expanded eligibility for endovascular thrombectomy to patients with large infarct cores and strokes of unclear onset. High-resolution vessel wall MRI and prolonged cardiac monitoring facilitate improved detection of high-risk atherosclerotic plaques and covert cardioembolic sources. Artificial intelligence increasingly supports acute stroke care by accelerating imaging interpretation, streamlining workflow, and monitoring posttreatment complications. Moreover, tenecteplase has emerged as a practical alternative to alteplase, offering comparable safety and efficacy with simplified single-bolus administration. Early initiation of dual antiplatelet therapy and lipid-lowering agents, including proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors, further suggests new opportunities to reduce recurrence and stabilize vulnerable plaques even during the acute phase of stroke management.

cal scales for triage, but these tools are limited in distinguishing ischemic from hemorrhagic stroke and in identifying the underlying stroke etiology [2]. Recent innovations, such as mobile stroke units with onboard imaging, point-of-care testing, advanced imaging protocols, and biomarker-based assays, now enable faster and more accurate stroke classification [3]. Prolonged cardiac monitoring and high-resolution vascular imaging further improve detection of cardioembolic sources and high-risk atherosclerotic plaques, helping identify the cause of ischemic stroke [4]. Artificial intelligence (AI) has extended its role beyond image interpretation to real-time clinical decision support, guiding eligibility for reperfusion therapies including intravenous thrombolysis (IVT) and endovascular thrombectomy (EVT).

Therapeutic advances are reshaping AIS care in the ED. New thrombolytic agents offer simplified administration with comparable or improved outcomes relative to current standards [5]. EVT indications have broadened to include large core, distal vessel, and posterior circulation strokes, guided by advanced imaging criteria and extended treatment windows [6]. Early implementation of adjunctive pharmacologic interventions, such as short-term dual antiplatelet therapy and early lipid-lowering therapy, may further decrease recurrence risk and enhance long-term vascular outcomes [7]. This review synthesizes recent developments in diagnosis and treatment, providing emergency physicians with practical, evidence-based insights to optimize stroke care from arrival to reperfusion.

PREHOSPITAL DIAGNOSIS: FROM TRIAGE TO ON-SITE DETECTION

Early identification of AIS in the prehospital setting has traditionally relied on clinical scoring systems such as the Cincinnati Prehospital Stroke Scale (CPSS) or the Los Angeles Prehospital Stroke Screen (LAPSS), as well as LVO-focused tools like the Rapid Arterial Occlusion Evaluation (RACE) [8–10]. These scales aid in recognizing stroke symptoms and estimating severity, facilitating transport to EVT-capable centers. However, their ability to distinguish ischemic from hemorrhagic stroke, the most critical determinant for initiating appropriate acute-phase reperfusion therapy, remains limited, and variability in provider training, patient presentation, and scale selection affects reliability [11].

Prehospital stroke care is increasingly shifting toward direct field diagnostics (Table 1) [3,8–21]. Mobile stroke units (MSUs), equipped with onboard computed tomography (CT) scanners, point-of-care laboratories, and telemedicine connectivity, enable imaging and treatment initiation before hospital arrival [3]. Randomized and observational studies have shown that MSUs reduce onset-to-needle times by 20 to 40 minutes and increase the proportion of patients treated within the "golden hour." Concurrently, blood-based biomarkers such as glial fibrillary acidic protein (GFAP) show promise for differentiating intracerebral hemorrhage from AIS both in prehospital environments (e.g., ambulances or MSUs) and in the ED. Point-of-care GFAP assays can provide re-

Table 1. Rapid diagnosis and triage in the ED

Tool	Implication for ED practice	Key evidence	Consideration
Mobile stroke unit	CT/POC/telemedicine before arrival → Faster IVT/EVT	Onset-to-needle ↓ 20–40 min [3] ↑ Golden-hour treatment [3]	Limited availability
GFAP POC test	Rapid ICH vs. AIS differentiation in prehospital and ED	Results in minutes [11,12] Improves triage accuracy [11,12]	Still needs imaging confirmation
Clinical scale (CPSS, LAPSS, RACE)	Quick triage Identify LVO for EVT referral	Simple, widely used [8–10]	Cannot distinguish ICH vs. AIS
Initial CT/CTA/CTP (±AI support)	Identify large core, LVO, collaterals	Large-core patients benefit from EVT [13–16] AI shortens door-to-puncture time [18–21]	CTP may overestimate core <3 hr
DWI-FLAIR mismatch (tissue clock)	Expands IVT to wake-up/unclear-onset stroke	WAKE-UP trial: mRS 0–1 at 90 days, 53.3% vs. 41.8% [17]	Requires MRI

ED, emergency department; CT, computed tomography; POC, point of care; IVT, intravenous thrombolysis; EVT, endovascular thrombectomy; GFAP, glial fibrillary acidic protein; ICH, intracerebral hemorrhage; AIS, acute ischemic stroke; CPSS, Cincinnati Prehospital Stroke Scale; LAPSS, Los Angeles Prehospital Stroke Screen; RACE, Rapid Arterial Occlusion Evaluation; LVO, large vessel occlusion; CTA, computed tomography angiography; CTP, computed tomography perfusion; AI, artificial intelligence; DWI, diffusion-weighted imaging; FLAIR, fluid-attenuated inversion recovery; mRS, modified Rankin scale; MRI, magnetic resonance imaging.

sults within minutes, refining thrombolysis eligibility and guiding transport strategies [12]. Together, these innovations may transition prehospital AIS care from symptom-based triage to rapid, on-site diagnostic confirmation. Advances in field-based diagnostics enable critical treatment decisions to be made immediately upon ED arrival. However, without imaging confirmation, precise treatment eligibility remains uncertain, underscoring the importance of next-generation stroke imaging strategies to expand EVT opportunities.

INITIAL IMAGING TO EXPAND EVT ELIGIBILITY

Recent advances in AIS imaging have focused on identifying large-core infarcts using CT and applying tissue clock concepts to select patients for EVT beyond conventional time windows [13,14]. The RESCUE-Japan LIMIT, ANGEL-ASPECT, and SELECT2 trials demonstrated that patients with an Alberta Stroke Program Early CT Score (ASPECTS) of 3–5 or a CT perfusion (CTP) core volume of ≥50–70 mL can still benefit from EVT, marking a shift away from excluding large-core patients [13,15,16]. Simple CT metrics such as ASPECTS and CTP-derived core estimates are now central to patient selection, while AI-driven alert systems in hub-and-spoke networks have been shown to shorten door-to-puncture times. Because CTP may overestimate core size within the first 3 hours, integrating ASPECTS, clinical presentation, and collateral status from single-phase or multiphase CT angiography (CTA) remains important.

For strokes with unclear onset, particularly wake-up strokes, imaging surrogates such as the diffusion-weighted imaging–fluid-attenuated inversion recovery mismatch can approximate onset time. The WAKE-UP trial and subsequent pooled analyses showed that this mismatch corresponds to a lesion age of about 3 hours, with more than 80% of cases occurring within 4.5 hours, thereby enabling safe IVT in patients previously considered ineligible. In the WAKE-UP trial, a modified Rankin scale (mRS) score of 0–1 at 90 days was achieved in 53.3% of patients compared with 41.8% in the placebo group [17]. Thus, imaging technologies are extending eligibility for both IVT and EVT beyond traditional temporal limits. Refinement of patient selection also enhances the identification of stroke etiology, particularly in cases of undetermined origin, which is crucial for guiding optimal secondary prevention strategies.

FINDING THE CULPRIT FOR STROKE

In patients with embolic stroke of undetermined source (ESUS), large randomized trials such as NAVIGATE ESUS and RE-SPECT ESUS failed to demonstrate the superiority of direct oral anticoagulants over aspirin for secondary prevention [22,23]. These findings have redirected attention toward identifying covert cardioembolic sources, particularly paroxysmal atrial fibrillation (AF). Prolonged cardiac monitoring significantly increases AF detection: the EMBRACE trial showed that 30-day external event monitoring detected AF in 16.1% of patients compared with 3.2% using 24-hour Holter monitoring, while the CRYSTAL-AF study found that implantable loop recorders identified AF in 12.4% of patients at 12 months, rising to 30% at 3 years [24,25]. Patch-type continuous electrocardiogram monitors now provide a less invasive, patient-friendly alternative to implantable loop recorders, maintaining high diagnostic yield over several weeks. Ongoing prospective trials, such as the Korean SOLO-ESUS study, are evaluating the clinical utility and cost-effectiveness of these

devices in guiding secondary prevention in ESUS [26].

High-resolution vessel wall magnetic resonance imaging (VW-MRI) is increasingly vital for characterizing intracranial atherosclerosis (ICAS) and identifying high-risk plaques associated with recurrent ischemic events. Beyond assessing luminal stenosis, VW-MRI can reveal intraplaque hemorrhage, lipid-rich necrotic core, irregular surfaces, and strong post-contrast enhancement [27]. In ICAS, culprit plaques with enhancement are correlated with higher recurrence rates and artery-to-artery embolism, suggesting that plaque inflammation and instability contribute to embolic events even in the absence of severe stenosis. The 4D flow-MRI studies have demonstrated that high wall shear stress is associated with plaque enhancement [28]. Vascular tortuosity has also been linked to ICAS development and altered local hemodynamics, potentially increasing plaque vulnerability. These findings underscore the importance of integrating structural and hemodynamic evaluations—via VW-MRI and flow modeling—for improved risk stratification and targeted prevention in ICAS and other intracranial vasculopathies.

USE OF AI IN THE ACUTE STAGE OF STROKE

AI has evolved from a tool for image postprocessing to a comprehensive decision-support system in AIS. The Rapid platform (iSchemaView Inc), one of the first widely adopted solutions, provides automated CT perfusion and MRI diffusion/perfusion analyses to rapidly quantify infarct core and penumbra. Its application in late-window thrombectomy trials such as DAWN and DEFUSE 3 demonstrated how automated imaging can standardize patient selection and accelerate treatment decisions across institutions [18,19]. More recently, AI has been applied to strokes of unclear onset, with models trained on imaging and clinical features predicting onset within 6 hours with greater than 80% accuracy, complementing tissue clock–based selection methods [29].

In Korea, AI platforms now perform fully automated LVO detection on CTA, ASPECTS scoring on noncontrast CT, and core-penumbra estimation on CT perfusion, all with processing times of only a few minutes [20]. Multicenter studies have shown high concordance with expert neuroradiologists (intraclass correlation coefficients of 0.6–0.8 for ASPECTS, and sensitivity for LVO exceeding 90%), as well as measurable workflow improvements in door-to-puncture times when AI triage alerts are used [21]. Some systems also extend to posttreatment monitoring, detecting hemorrhagic transformation or malignant edema on follow-up imaging. These developments illustrate a transition toward AI as a platform not only for patient selection but also for therapy prioritization, complication monitoring, and quality improvement. With expanding regulatory approvals and integration into clinical workflows, AI-based decision support is poised to become a standard component of acute stroke care in Korea and globally. Acute treatment strategies and early secondary prevention approaches for AIS are summarized in Table 2 [5,13,15–19,30–43].

Tenecteplase in AIS

Tenecteplase, a genetically modified variant of alteplase with higher fibrin specificity and a longer half-life, enables single-bolus administration, thereby simplifying workflow and potentially reducing door-to-needle time [30–34] (Table 2). Multiple randomized controlled trials have compared tenecteplase with standard-dose alteplase. In the EXTEND-IA TNK trial, patients eligible for EVT who received tenecteplase 0.25 mg/kg achieved higher pre-procedure reperfusion rates (22% vs. 10%, P = 0.002) and better 90-day functional outcomes, without an increase in symptomatic intracranial hemorrhage [30]. The ATTEST trial demonstrated similar early neurological improvement and safety [31], with a trend toward greater reperfusion in the tenecteplase group [32].

Large-scale studies have reinforced these findings in broader AIS populations. The AcT trial, involving 1,577 patients within 4.5 hours of onset, showed that tenecteplase 0.25 mg/kg was noninferior to alteplase in achieving an mRS score of 0–1 at 90 days (36.9% vs. 34.8%; risk difference, 2.1%; 95% confidence interval [CI], −2.6 to 6.9) with no difference in symptomatic hemorrhage [33]. The TRACE-2 trial in China confirmed noninferiority in an Asian cohort [34]. Meta-analyses indicate that tenecteplase offers at least equivalent functional outcomes, with potential advantages in reperfusion and operational efficiency [44,45].

Given its ease of administration, favorable workflow impact, and comparable safety profile, international guidelines increasingly recognize tenecteplase, particularly at a dose of 0.25 mg/kg, as a practical alternative to alteplase in eligible AIS patients. Ongoing studies are investigating its use in extended time windows, wake-up strokes, and other subgroups, potentially further expanding its role in acute reperfusion therapy.

EVT in AIS

EVT for large ischemic cores is now supported by multiple randomized controlled trials [13,15,16] (Table 2). In RESCUE-Japan LIMIT (ASPECTS 3–5 or MRI-guided ≤24 hours), an mRS of 0–3 at 90 days was achieved in 31.0% of EVT patients compared with 12.7% receiving medical therapy (risk ratio [RR], 2.43), although

Table 2. Acute treatment and early secondary prevention

Therapy	Implication for ED practice	Benefit	Risk/consideration
Alteplase (tPA)	0.9 mg/kg infusion ≤4.5 hr	Standard treatment [5,17]	Infusion time, monitoring
Tenecteplase	0.25 mg/kg single bolus ≤4.5 hr	Higher fibrin specificity; longer half-life noninferior to tPA [30–34] EVT reperfusion 22% vs. 10% [30]	Not yet approved for AIS in Korea (regulatory review ongoing)
EVT			
Large core	ASPECTS 3–5, core ≥50–70 mL	Functional independence ↑ [13,15,16]	Higher ICH risk
Late/unknown onset	DAWN trial [18]: 6–24 hr DEFUSE 3 trial [19]: 6–16 hr	mRS 0–2: 49% vs. 13% [18] 45% vs. 17% [19]	Substantial residual disability
Posterior circulation occlusion	ATTENTION trial [35]: basilar artery ≤24 hr	mRS 0–3: 46% vs. 23% with medical care [35]	Prompt transfer essential due to rapid deterioration
Post-EVT BP target	Keep SBP 140–180 mmHg (avoid <140 mmHg early)	OPTIMAL-BP trial [36]: better outcomes with standard vs. intensive	No benefit and potential harm from intensive lowering
Short-term DAPT			
Aspirin+Clopidogrel	Minor stroke/TIA	Reduced recurrent stroke CHANCE trial [37]: 8.2% vs. 11.7% POINT trial [38]: 5.0% vs. 6.5% INSPIRES trial [39]: 7.3% vs. 9.2%	Bleeding ↑ if >21–30 days
Aspirin+Ticagrelor	Alternative DAPT option, faster onset, CYP2C19-independent	Lowered 30-day stroke/death THALES trial [40]: 5.5% vs. 6.6%	Severe bleeding ↑ (0.5% vs. 0.1%)
High-intensity statin (±PCSK9 inhibitor)	Initiate in ED	Rapid LDL-C ↓; plaque stabilization [41–43]	No large RCT on short-term neurological outcomes

ED, emergency department; tPA, tissue plasminogen activator; EVT, endovascular thrombectomy; AIS, acute ischemic stroke; ASPECTS, Alberta Stroke Program Early CT Score; ICH, intracerebral hemorrhage; mRS, modified Rankin scale; BP, blood pressure; SBP, systolic blood pressure; DAPT, dual antiplatelet therapy; TIA, transient ischemic attack; CYP2C19, cytochrome P450 2C19; PCSK9, proprotein convertase subtilisin/kexin type 9; LDL-C, low-density lipoprotein cholesterol; RCT, randomized controlled trial.

any intracranial hemorrhage occurred more frequently in the EVT group (58.0% vs. 31.4%) [15]. ANGEL-ASPECT (ASPECTS 3–5 or core 70–100 mL; ≤24 hours) demonstrated a favorable shift in mRS distribution (generalized odds ratio [OR], 1.37; 95% CI, 1.11–1.69), with symptomatic intracranial hemorrhage in 6.1% vs. 2.7% [16]. SELECT2 (ASPECTS 3–5 or core ≥50 mL; ≤24 hours) was terminated early for efficacy, showing an mRS shift OR of 1.51 (95% CI, 1.20–1.89) and functional independence in 20% vs. 7% (RR, 2.97), with no increase in mortality [13].

For unclear-onset and late-window strokes, imaging-based selection remains central. In DAWN (last known well 6–24 hours; clinical-core mismatch), functional independence was achieved in 49% vs. 13%, with an mRS shift of 5.5 vs. 3.4. DEFUSE 3 (6–16 hours; CTP mismatch) showed an mRS 0–2 in 45% vs. 17%, and mortality was lower (14% vs. 26%). Nevertheless, pooled late-window data reveal substantial residual disability: in DEFUSE 3, 39% had an mRS ≥4, and in DAWN, 42%, underscoring the need for careful preprocedural counseling. EVT for posterior circulation occlusion has also expanded: the ATTENTION trial (basilar artery ≤24 hours) achieved mRS 0–3 in 46% vs. 23% with medical management, supporting EVT beyond conventional limits in selected patients [18,19,35].

Blood pressure (BP) control after EVT is crucial to prevent reperfusion injury, hemorrhagic transformation, and infarct expansion. Higher postprocedural systolic BP and greater BP variability have been linked to poorer outcomes and higher symptomatic intracranial hemorrhage risk. The OPTIMAL-BP trial randomized patients after successful endovascular thrombectomy to intensive systolic BP lowering (<140 mmHg) versus a standard target (140–180 mmHg) within 1 hour of recanalization. At 90 days, functional independence (mRS ≤2) was lower in the intensive group (39.4%) than in the standard group (54.4%), suggesting no benefit—and possible harm—from aggressive BP reduction [36] (Table 2).

ANTIPLATELET AND LIPID MANAGEMENT IN THE ACUTE STAGE

Dual antiplatelet therapy (DAPT) with aspirin and clopidogrel is supported for acute minor ischemic stroke and high-risk transient ischemic attack. In CHANCE trial (n=5,170), 21-day DAPT reduced 90-day recurrent stroke (8.2% vs. 11.7%; hazard ratio [HR], 0.68) without an increase in major bleeding; subgroup analyses showed greater benefit in patients with large-artery atherosclerosis and those treated within 12 hours of onset [37]. The POINT trial (n=4,881) confirmed efficacy (5.0% vs. 6.5%; HR, 0.75) but

showed higher rates of major hemorrhage (0.9% vs. 0.4%); benefits were concentrated in the first 21 days [38]. The INSPIRES trial (n = 6,100), enrolling patients with ≥ 50% stenosis, showed reduced new stroke (7.3% vs. 9.2%; HR, 0.79) but higher moderate-to-severe bleeding (0.9% vs. 0.4%); effects were consistent across intracranial and extracranial disease [39]. The THALES trial (n = 11,016) found that aspirin-ticagrelor reduced 30-day stroke or death (5.5% vs. 6.6%; HR, 0.83), with the greatest benefit in ipsilateral large-artery atherosclerosis but a higher rate of severe bleeding (0.5% vs. 0.1%) [40] (Table 2).

Proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors, including evolocumab and alirocumab, have demonstrated in coronary artery disease that early initiation can rapidly lower low-density lipoprotein cholesterol and promote plaque stabilization, as evidenced by intravascular imaging studies such as GLAGOV and PACMAN-AMI [41,42] (Table 2). In ischemic stroke, particularly in Chinese cohorts, small prospective studies have investigated early post-AIS use, reporting significant low-density lipoprotein cholesterol reduction, improved plaque characteristics on vessel wall MRI [43], and favorable modulation of inflammatory markers within weeks. While these findings suggest a potential role in secondary prevention beginning in the acute phase, no large randomized trials have yet confirmed benefits on short-term neurological outcomes. Further studies are needed to determine whether early PCSK9 inhibitor therapy after AIS can reproduce the plaque-stabilizing and event-reducing effects observed in acute coronary syndrome.

CONCLUSION

The landscape of AIS care is rapidly evolving, with advances in prehospital diagnostics, imaging technologies, reperfusion therapies, and pharmacologic strategies expanding treatment opportunities and improving clinical outcomes. The integration of MSUs, point-of-care biomarkers, and AI-driven decision-support systems represents a shift toward rapid, mechanism-informed intervention. Moving forward, the successful translation of these innovations into routine clinical practice will depend not only on robust supporting evidence but also on seamless, multidisciplinary collaboration among emergency medicine, neurology, radiology, neurosurgery, and rehabilitation teams to ensure timely, coordinated, and comprehensive stroke care.

ARTICLE INFORMATION

Author contributions

Conceptualization: SHH, SK, BJK; Investigation: SHH, SK, BJK; Methodology: SMK, JP, SEL, BJK; Resources: SMK, JP, SEL, BJK; Supervision: SMK, JP, SEL, BJK; Visualization: SMK, JP, SEL, BJK; Writing–original draft: SHH, SK, BJK; Writing–review & editing: all authors. All authors read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

Funding

The authors received no financial support for this study.

Data availability

Data sharing is not applicable as no new data were created or analyzed in this study.

REFERENCES

1. Saver JL. Time is brain: quantified. Stroke 2006;37:263–6.
2. Duvekot MH, Venema E, Lingsma HF, et al. Sensitivity of prehospital stroke scales for different intracranial large vessel occlusion locations. Eur Stroke J 2021;6:194–204.
3. Röhrs KJ, Audebert H. Pre-hospital stroke care beyond the MSU. Curr Neurol Neurosci Rep 2024;24:315–22.
4. Fleg JL, Stone GW, Fayad ZA, et al. Detection of high-risk atherosclerotic plaque: report of the NHLBI working group on current status and future directions. JACC Cardiovasc Imaging 2012;5:941–55.
5. Lin J, Zuo W, Jin H, et al. Thrombolysis for acute ischaemic stroke: development and update. Brain Commun 2025;7:fcaf164.
6. Ha SH, Ryu JC, Bae JH, et al. Early response to endovascular thrombectomy after stroke: early, late, and very late time windows. Cerebrovasc Dis 2023;52:28–35.
7. Gao Y, Pan Y, Han S, et al. Rationale and design of a randomised double-blind 2 × 2 factorial trial comparing the effect of a 3-month intensive statin and antiplatelet therapy for patients with acute mild ischaemic stroke or high-risk TIA with intracranial or extracranial atherosclerosis (INSPIRES). Stroke Vasc Neurol 2023;8:249–58.
8. Kothari RU, Pancioli A, Liu T, Brott T, Broderick J. Cincinnati prehospital stroke scale: reproducibility and validity. Ann Emerg Med 1999;33:373–8.

9. Kidwell CS, Starkman S, Eckstein M, Weems K, Saver JL. Identifying stroke in the field: prospective validation of the Los Angeles Prehospital Stroke Screen (LAPSS). Stroke 2000;31:71–6.
10. Pérez de la Ossa N, Carrera D, Gorchs M, et al. Design and validation of a prehospital stroke scale to predict large arterial occlusion: the rapid arterial occlusion evaluation scale. Stroke 2014;45:87–91.
11. Almubayyidh M, Alghamdi I, Jenkins D, Parry-Jones A. Systematic review of prehospital prediction models for identifying intracerebral haemorrhage in suspected stroke patients. Healthcare (Basel) 2025;13:876.
12. Almubayyidh M, Jenkins DA, Gaude E, Parry-Jones AR. Integrating clinical predictors and glial fibrillary acidic protein in prediction models for the prehospital identification of intracerebral haemorrhage in suspected stroke. BMJ Neurol Open 2025;7:e001160.
13. Sarraj A, Hassan AE, Abraham MG, et al. Trial of endovascular thrombectomy for large ischemic strokes. N Engl J Med 2023;388:1259–71.
14. Olthuis SG, Pirson FA, Pinckaers FM, et al. Endovascular treatment versus no endovascular treatment after 6-24 h in patients with ischaemic stroke and collateral flow on CT angiography (MR CLEAN-LATE) in the Netherlands: a multicentre, open-label, blinded-endpoint, randomised, controlled, phase 3 trial. Lancet 2023;401:1371–80.
15. Yoshimura S, Sakai N, Yamagami H, et al. Endovascular therapy for acute stroke with a large ischemic region. N Engl J Med 2022;386:1303–13.
16. Huo X, Ma G, Tong X, et al. Trial of endovascular therapy for acute ischemic stroke with large infarct. N Engl J Med 2023;388:1272–83.
17. Thomalla G, Boutitie F, Ma H, et al. Intravenous alteplase for stroke with unknown time of onset guided by advanced imaging: systematic review and meta-analysis of individual patient data. Lancet 2020;396:1574–84.
18. Nogueira RG, Jadhav AP, Haussen DC, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. N Engl J Med 2018;378:11–21.
19. Albers GW, Marks MP, Kemp S, et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. N Engl J Med 2018;378:708–18.
20. Lee SJ, Park G, Kim D, et al. Clinical evaluation of a deep-learning model for automatic scoring of the Alberta stroke program early CT score on non-contrast CT. J Neurointerv Surg 2023;16:61–6.
21. Wei J, Shang K, Wei X, et al. Deep learning-based automatic ASPECTS calculation can improve diagnosis efficiency in patients with acute ischemic stroke: a multicenter study. Eur Radiol 2025;35:627–39.
22. Hart RG, Sharma M, Mundl H, et al. Rivaroxaban for stroke prevention after embolic stroke of undetermined source. N Engl J Med 2018;378:2191–201.
23. Diener HC, Sacco RL, Easton JD, et al. Dabigatran for prevention of stroke after embolic stroke of undetermined source. N Engl J Med 2019;380:1906–17.
24. PérezRodon J, FranciscoPascual J, RivasGándara N, RocaLuque I, Bellera N, MoyaMitjans À. Cryptogenic stroke and role of loop recorder. J Atr Fibrillation 2014;7:1178.
25. De Angelis G, Cimon K, Sinclair A, et al. Monitoring for atrial fibrillation in discharged stroke and transient ischemic attack patients: a clinical and cost-effectiveness analysis and review of patient preferences. CADTH Optimal Use Report, No. 5.2b. Canadian Agency for Drugs and Technologies in Health; 2016.
26. Song K, Song TJ, Woo HG, et al. Rationale and design of the efficacy and safety of a patch-type cardiac monitor for diagnosing paroxysmal atrial fibrillation in embolic stroke of undetermined source patients with left atrial enlargement (SOLO-ESUS): a randomized controlled trial. Cerebrovasc Dis 2026;55:83–90.
27. Woo HG, Kim HG, Lee KM, et al. Wall shear stress associated with stroke occurrence and mechanisms in middle cerebral artery atherosclerosis. J Stroke 2023;25:132–40.
28. Sakai Y, Lehman VT, Eisenmenger LB, et al. Vessel wall MR imaging of aortic arch, cervical carotid and intracranial arteries in patients with embolic stroke of undetermined source: a narrative review. Front Neurol 2022;13:968390.
29. Wouters A, Dupont P, Christensen S, et al. Multimodal magnetic resonance imaging to identify stroke onset within 6 h in patients with large vessel occlusions. Eur Stroke J 2018;3:185–92.
30. Campbell BCV, Mitchell PJ, Churilov L, et al. Tenecteplase versus alteplase before thrombectomy for ischemic stroke. N Engl J Med 2018;378:1573–82.
31. Huang X, Cheripelli BK, Lloyd SM, et al. Alteplase versus tenecteplase for thrombolysis after ischaemic stroke (ATTEST): a phase 2, randomised, open-label, blinded endpoint study. Lancet Neurol 2015;14:368–76.
32. Parsons M, Spratt N, Bivard A, et al. A randomized trial of tenecteplase versus alteplase for acute ischemic stroke. N Engl J Med 2012;366:1099–107.
33. Menon BK, Buck BH, Singh N, et al. Intravenous tenecteplase

compared with alteplase for acute ischaemic stroke in Canada (AcT): a pragmatic, multicentre, open-label, registry-linked, randomised, controlled, non-inferiority trial. Lancet 2022; 400:161–9.
34. Wang Y, Li S, Pan Y, et al. Tenecteplase versus alteplase in acute ischaemic cerebrovascular events (TRACE-2): a phase 3, multicentre, open-label, randomised controlled, non-inferiority trial. Lancet 2023;401:645–54.
35. Tao C, Nogueira RG, Zhu Y, et al. Trial of endovascular treatment of acute basilar-artery occlusion. N Engl J Med 2022; 387:1361–72.
36. Nam HS, Kim YD, Choi JK, et al. Outcome in patients treated with intra-arterial thrombectomy: the optiMAL Blood Pressure control (OPTIMAL-BP) trial. Int J Stroke 2022;17:931–7.
37. Wang Y, Wang Y, Zhao X, et al. Clopidogrel with aspirin in acute minor stroke or transient ischemic attack. N Engl J Med 2013;369:11–9.
38. Johnston SC, Easton JD, Farrant M, et al. Clopidogrel and aspirin in acute ischemic stroke and high-risk TIA. N Engl J Med 2018;379:215–25.
39. Gao Y, Chen W, Pan Y, et al. Dual antiplatelet treatment up to 72 hours after ischemic stroke. N Engl J Med 2023;389:2413–24.
40. Johnston SC, Amarenco P, Denison H, et al. Ticagrelor and aspirin or aspirin alone in acute ischemic stroke or TIA. N Engl J Med 2020;383:207–17.
41. Nicholls SJ, Puri R, Anderson T, et al. Effect of evolocumab on progression of coronary disease in statin-treated patients: the GLAGOV randomized clinical trial. JAMA 2016;316:2373–84.
42. Räber L, Ueki Y, Otsuka T, et al. Effect of alirocumab added to high-intensity statin therapy on coronary atherosclerosis in patients with acute myocardial infarction: the PACMAN-AMI randomized clinical trial. JAMA 2022;327:1771–81.
43. Zeng W, Zhou F, Zhao H, et al. Evaluation of intensive statins and proprotein convertase subtilisin/kexin type 9 inhibitors on intracranial artery plaque stability: a prospective single-arm study. J Am Heart Assoc 2025;14:e035651.
44. Abuelazm M, Seri AR, Awad AK, et al. The efficacy and safety of tenecteplase versus alteplase for acute ischemic stroke: an updated systematic review, pairwise, and network meta-analysis of randomized controlled trials. J Thromb Thrombolysis 2023;55:322–38.
45. Kobeissi H, Ghozy S, Turfe B, et al. Tenecteplase vs. alteplase for treatment of acute ischemic stroke: a systematic review and meta-analysis of randomized trials. Front Neurol 2023; 14:1102463.

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):13-27
https://doi.org/10.15441/ceem.25.059

Review Article

The role of fomepizole in acetaminophen-related poisoning: a narrative review

Frank Chen[1], Shelly ZQ Lu[2], Harleen Chohan[2,3], Anthony Lau[2,4]

[1]Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
[2]Department of Pharmacy, Vancouver General Hospital, Vancouver, BC, Canada
[3]Solid Organ Transplant Program, Vancouver General Hospital, Vancouver, BC, Canada
[4]Department of Emergency Medicine, Vancouver General Hospital, Vancouver, BC, Canada

pISSN: 3022-1811
eISSN: 2383-4625

N-acetylcysteine (NAC) as the gold standard antidote in cases of acetaminophen (APAP)-related poisoning might not be sufficient with delayed presentation or massive ingestion. Human reports published up to July 2021 suggest that fomepizole could play a role in APAP overdoses by inhibiting cytochrome P450 2E1 (CYP2E1)-mediated N-acetyl-p-benzoquinone imine (NAPQI) production and JNK-mediated oxidative damage. This narrative review builds on previous systematic and scoping reviews by adding the latest evidence about the use of fomepizole in APAP poisoning to better understand the hepatoprotective role and safety profile of this medication, as well as its practical place in therapy. A systematic search of MEDLINE and Embase was completed through November 2024. Studies involving human patients with APAP toxicity who received fomepizole treatment were included. Each case was summarized in tables to identify clinical trends, particularly the risk of hepatotoxicity, quantity of ingestion, time of presentation since ingestion, therapeutic and dosing regimens, and clinical outcomes. This review covers 28 studies and 45 patients across 18 case reports and 6 case series. When used with NAC, fomepizole seemed to result in favorable laboratory and clinical outcomes in most patients at high risk of hepatotoxicity because of late presentation or massive APAP ingestion. The available data suggest that fomepizole might complement NAC in treating severe APAP toxicity. Though they lack detailed clinical outcome analyses, case studies suggest that fomepizole could improve hepatotoxicity, survival, and transplant-free days.

Keywords: Acetaminophen; Antidotes; Fomepizole; Liver; Toxicology

Capsule Summary

What is already known

N-acetylcysteine as the gold standard antidote may not be sufficient in managing cases of acetaminophen (APAP)-related poisoning with delayed presentations or with massive ingestions. Previous scoping and systematic reviews of human data up until July 2021 have suggested that fomepizole may play a potential role in APAP overdoses through inhibition of cytochrome P450 2E1 (CYP2E1)-mediated N-acetyl-p-benzoquinone imine (NAPQI) and JNK-mediated oxidative damage, but they were unclear regarding the hepatoprotective benefits of fomepizole in treating APAP poisoning.

What is new in the current study

This narrative review summarizes the most recent evidence about using fomepizole in humans who have overdosed on APAP to offer new insights into its safety, effectiveness, and role in therapy.

Received: March 11, 2025
Revised: April 29, 2025
Accepted: May 7, 2025

Correspondence to:
Anthony Lau
(Anthony.Lau@vch.ca)

How to cite this article:
Chen F, Lu SZQ, Chohan H, Lau A. The role of fomepizole in acetaminophen-related poisoning: a narrative review. Clin Exp Emerg Med 2026;13(1):13-27. https://doi.org/10.15441/ceem.25.059

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

© 2026 The Korean Society of Emergency Medicine

INTRODUCTION

Canada faces nearly 4,500 hospitalizations each year from acetaminophen (APAP)-related poisoning, with 6% of them resulting in liver injuries and acute liver failure requiring liver transplants or leading to death [1]. At therapeutic doses, 50% to 70% of APAP is metabolized by glucuronidation, 25% to 35% by sulfation, and 2% to 5% is excreted unchanged in urine [2]. The remainder undergoes liver metabolism, primarily by cytochrome P450 2E1 (CYP2E1), which generates toxic N-acetyl-p-benzoquinone imine (NAPQI), a key driver of APAP-induced hepatotoxicity [2]. At toxic doses (>4 g/day in adults or >75 mg/kg/day in children), the sulfation process becomes saturated, leading to excessive NAPQI production that depletes intracellular glutathione, causing oxidative liver damage. Fomepizole protects the liver by inhibiting CYP2E1, which reduces NAPQI formation, and blocking JNK activation to mitigate mitochondrial oxidative stress [2].

As the only available antidote for APAP-related poisoning, N-acetylcysteine (NAC) is most effective if initiated within 8 to 10 hours after an overdose, and it is successful in managing most toxic cases. However, with a massive ingestion APAP levels above the 300 μg line on the Rumack-Matthew Nomogram or >30 g of ingestion or late presentation (>8–10 hours), NAC alone might not prevent the need for liver transplantation or fulminant liver failure, prompting inquiries into whether additional therapy could be provided [2].

In the absence of robust studies, changes in standards of practice rely heavily on the body of evidence available in human case reports and case series. Previous systematic or scoping reviews summarized preclinical *in vivo* and *in vitro* data, along with human case studies on fomepizole in APAP-related poisoning only up to July 2021, and they offered preliminary and inconclusive insights into its use. This narrative review focuses on human experiences and includes several new case studies to identify emerging trends and patterns and better inform more specific recommendations for the indications and role of fomepizole in managing APAP poisoning.

METHODS

Study design

This review summarizes the current evidence for the use of fomepizole to manage APAP overdose and toxicity. A systematic literature search was performed in the MEDLINE and Embase databases from their inception to November 2024 to identify relevant studies on this topic. The search strategy, developed with the assistance of a librarian, included Medical Subject Headings (MeSH) terms such as "fomepizole," "acetaminophen," "antidote," "overdose," and "toxicity" (Suppl. 1).

Selection of studies

The review included full-text studies and abstracts from conference proceedings that were written in English and involved the use of fomepizole as an antidote for patients with APAP overdoses. Animal studies, non-English publications, and studies exploring other indications of fomepizole, such as toxic alcohol ingestion in the absence of APAP overdose, were excluded. The references of reviews about fomepizole therapy for APAP toxicity management were hand-searched to identify any original articles that might otherwise have been missed. Two independent reviewers screened titles and abstracts, with a third resolving conflicts and discrepancies, and the data were extracted using a standardized form. Each article was thoroughly screened, and the study design, patient characteristics, treatment protocols, and key outcomes (recovery in liver function or injury, acid-base status, mental status, and timing of discharge) were extracted and summarized in tables. The risk of bias in each included study was identified using a full-text review and the JBI critical appraisal tool (Suppl. 2, 3).

Study analysis

A narrative synthesis was performed by identifying key themes that highlight human experiences with fomepizole protocols, including dosing, concomitant use with NAC and dialysis, timing of initiation, place in therapy (i.e., massive ingestion and late presentation), and effects on patient outcomes based on the risk of hepatotoxicity. No statistical analyses of the results were possible because of the large number of case series and heterogeneity in the data found.

RESULTS

The search strategy yielded 214 potentially relevant citations. After the removal of duplicate citations and initial screening, 51 studies were selected for full article review. Twenty-three studies were excluded after the full-text article review. Therefore, 28 publications were included, comprising 18 case reports and 6 case series describing 45 subjects, along with 1 randomized crossover trial, 1 registry-based study, and 2 conference abstracts [3–30]. One patient was excluded from a case series by Link et al. [21] due to duplication because the same case had already been reported by the same authors in 2020 [15]. Tables 1 and 2 provide

Table 1. Summary of included case reports about fomepizole use in acetaminophen overdose

Study	Age (yr)	Sex	Medical history, presentation, timing	Ingestion dose and timing (if known)	Intervention and timing of interventions after ingestion unless stated otherwise (if known)	Laboratory finding[a]		Clinical course
						APAP level	Liver enzyme level	
Williams et al. [3] (2024)	18	Male	Depression, anxiety, borderline personality disorder Nausea, vomiting	APAP 125 g (1 hr prior)	4 hr: NAC 150 mg/kg IV then 12.5 mg/kg/hr IV × 3 days 5 hr: intubation, norepinephrine, vasopressin 9 hr: fomepizole 15 mg/kg IV, hemodialysis for 4 hr × 3 days 16 hr: CRRT for 4 hr, NAC 70 mg/kg PO every 4 hr, fomepizole decreased to 10 mg/kg IV every 6 hr	At 4 hr: 738 μg/mL (undetectable at 57 hr)	Day 2: AST, 50 U/L Day 4: ALT, 55 U/L	Developed shock, decline in mental status, acidosis at 5 hr Mental status, liver enzymes returned to baseline on day 3, and no permanent liver toxicities developed
Chary et al. [4] (2023)	50	Female	SUD Unresponsiveness, somnolence, ethanol 293 mg/dL	-	Preadmission: naloxone 4 mg IN and 2 mg IV Admission: NAC 150 mg/kg IV, then 12.5 mg/kg/hr IV for 4 hr, then 6.25 mg/kg/hr IV for 16 hr After NAC: intubation Day 4: fomepizole 15 mg/kg IV and cryoprecipitate 6 U 12 hr after Day 5: fomepizole 10 mg/kg IV, CRRT Day 6: filgrastim 300 μg	Days 1–3: 200 to > 530 μg/mL (< 100 μg/mL on day 5)	Days 3–4: AST, > 6,000 U/L (normalized on day 5) ALT, 6,600 U/L	Patient developed grade III encephalopathy and started CRRT on day 5 Developed neutropenia, and filgrastim was started on day 6 Mental status improved by day 7 and returned to baseline by day 10
Lobo et al. [5] (2023)	59	Male	Encephalopathy	APAP 100 g, ibuprofen 100 g (5 hr prior)	Timing unknown: NAC, fomepizole 15 mg/kg IV After transfer to tertiary facility: NAC 12.5 mg/kg/hr IV, fomepizole 10 mg/kg IV × 5 doses 31 hr: CRRT × 24 hr	At 24 hr: 535 μg/mL (undetectable at 105 hr)	Timing unknown: ALT, 123 U/L (normalized at 97 hr) AST, 100 U/L (normalized at 5 hr)	Patient discharged on day 25
Pepin et al. [6] (2023)	7 mo	Female	Fever, vomiting, cough, encephalopathy	APAP 4.8 g, ibuprofen (over 5 days)	Timing unknown: NAC 150 mg/kg IV over 1 hr, then 50 mg/kg IV over 4 hr, then 100 mg/kg IV over 16 hr 8 hr after admission: fomepizole 15 mg/kg IV After fomepizole: fresh frozen plasma, platelets, aminocaproic acid 500 mg, vitamin K 3 mg IV After transfer to transplant facility: fomepizole 10 mg/kg IV every 12 hr × 3 doses, NAC 12.5 mg/kg/hr IV × 81 hr	Timing unknown: 239.6 μg/mL (1 μg/mL on day 5)	8 hr after admission: AST, > 27,300 U/L (216 U/L on day 6) ALT, > 11,700 U/L (2,403 U/L on day 6)	Worsening agitation observed on day 4 NAC was stopped on day 4, and patient returned to mental baseline on day 5 and was discharged on day 8
Ho and Brambl [7] (2023)	37	Male	Tachycardia, hypothermia, tachypnoea, hypertension, acidemia, hepatotoxicity, possible alcohol ingestion	-	Timing unknown intubation, activated charcoal 50 g, NAC 140 mg/kg PO, NAC 150 mg/kg IV then 24 mg/kg/hr IV, fomepizole 15 mg/kg IV then 10 mg/kg IV every 12 hr × 4 doses 12 hr after admission: hemodialysis	On admission: 707 μg/mL (undetectable on day 3)	Day 2: transaminases, > 3,000 U/L	Despite fomepizole, charcoal, NAC, and hemodialysis, patient continued to be acidotic and developed significant hepatotoxicity Other outcomes were not documented

(Continued on the next page)

Table 1. (Continued)

Study	Age (yr)	Sex	Medical history, presentation, timing	Ingestion dose and timing (if known)	Intervention and timing of interventions after ingestion unless stated otherwise (if known)	Laboratory finding[a]		Clinical course
						APAP level	Liver enzyme level	
Kusnik et al. [8] (2022)	58	Female	Fibromyalgia, AUD Altered mental status, abdominal pain ×2 days, AKI	Unknown quantity (6–18 hr prior)	6–18 hr: NAC ×21 hr, multivitamins, pantoprazole 15–27 hr: NAC increased to 12.5 mg/kg/hr IV, then decreased to 6.25 mg/kg/hr IV after APAP undetectable 36–48 hr: fomepizole 15 mg/kg IV then 10 mg/kg IV every 12 hr given at 30, 45, 57, 69, and 81 hr Timing unknown: albumin, midodrine, octreotide	On 6–18 hr: 126.3 μg/mL (undetectable at 40 hr)	Days after 27 hr: AST, 34,665–13,995 U/L (normalized ≈4 wk after admission)	Kidney function improved 80 hr after admission without hemodialysis Encephalopathy improved 1 day after admission Liver and kidney function returned to baseline 4 wk after admission
Bassi et al. [9] (2022)	23	Male	Agitation, laryngospasms	APAP 188.5 g, naproxen, ibuprofen, hydrocodone, loratadine (1–3 hr prior)	Admission: NAC 200 mg/kg IV over 4 hr, then 100 mg/kg IV over 16 hr, activated charcoal 50g 34 hr: NAC infusion rate doubled, fomepizole 15 mg/kg IV, activated charcoal 50g 47 hr: endoscopy Before transfer to transplant facility: CRRT	20 hr after admission: 287 μg/mL (undetectable at 96 hr)	AST (timing unknown): 209 and 298 U/L	Recovered without liver transplant
Dion et al. [10] (2022)	17	Male	Undiagnosed eating disorder Abdominal pain, altered mental status, apnea	APAP 50–75 g, caffeinated drink, sennosides (2 hr prior)	3 hr: NAC 300 mg/kg IV, fomepizole 15 mg/kg IV Day 2: fomepizole 15 mg/kg IV, NAC 300 mg/kg IV ×48 hr	At 12 hr: 391.8 μg/mL (2.8 μg/mL at 60 hr)	No significant increases in liver enzymes were observed	Abdominal pain resolved by day 3, and patient was cleared for discharge on day 4
Chiu et al. [11] (2021)	20	Female	Anxiety, self-harm/prior suicide attempts Tachycardia, refractory shock, altered LOC, metabolic acidosis	APAP >120 g, ibuprofen 65 g, oxycodone 300 mg (2 hr prior)	Timing unknown: intubation, norepinephrine 0.3 μg/kg/min ×2 days, epinephrine 0.2 μg/kg/min ×2 days, sodium bicarbonate 3 hr: activated charcoal, NAC 150 mg/kg IV then 15 mg/kg/hr IV ×3 days, fomepizole 1,000 mg IV (10 mg/kg IV every 4 hr during dialysis and for an additional day 700 mg IV every 12 hr; total 2 days of fomepizole) 6 hr: hemodialysis	At 7 hr: ≈680 μg/mL (undetectable at 32 hr)	Admission day 2: ALT 76 U/L	24 hr after admission, patient was awake and oriented and followed commands Metabolic acidosis was refractory to NAC and fomepizole, requiring dialysis Acid-base/hemodynamic status returned to baseline without organ failure
Homan et al. [12] (2021)	58	Female	AUD, hypertension, hyperlipidemia, hepatic steatosis Abdominal pain, nausea, vomiting, lethargy, encephalopathy, RUQ tenderness	-	Timing unknown: fluids, NAC, fomepizole (unknown dose), lactulose, rifaximin, vitamin K, albumin, octreotide, midodrine	Unknown timing: 58.3 (units unknown)	Unknown timing: AST, 19,816 ALT, 5,544 (units unknown)	Pancreatitis, coagulopathy, and hepatorenal syndrome developed Mental status and laboratory abnormalities improved gradually, and APAP levels became undetectable

(Continued on the next page)

Table 1. (Continued)

Study	Age (yr)	Sex	Medical history, presentation, timing	Ingestion dose and timing (if known)	Intervention and timing of interventions after ingestion unless stated otherwise (if known)	Laboratory finding[a]		Clinical course
						APAP level	Liver enzyme level	
Cuninghame et al. [13] (2021)	24	Female	Endometriosis, chronic pelvic pain, depression, anxiety, previous suicide attempts Altered mental status, metabolic acidosis	APAP 50 g, gabapentin 25.5 g, morphine 60 mg (16 hr prior)	16 hr: NAC 60 mg/kg IV then 12 mg/kg/hr IV, fomepizole 600 mg IV q6 hr 18 hr: intubation 43 hr: hemodialysis 56 hr: second hemodialysis run 66 hr: CRRT 48 hr after CRRT: whole bowel irrigation, colonic decompression 16 days after admission: life-sustaining measures withdrawn	18 hr: 602 μg/mL (undetectable on admission day 8)	88 hr: transaminases > 8,000 U/L	Fulminant liver failure progressed with cerebral edema, encephalopathy, hypoglycemia, petechial hemorrhage, diffuse axonal injury, coagulopathy Altered mental status and metabolic acidosis were refractory to all interventions First case report of fatal APAP ingestion refractory to both hemodialysis and CRRT at time of writing
Woolum et al. [14] (2020)	55 M with	Male	Type I bipolar depressive disorder, previous suicidal ideation, diabetes, CKD stage III Unresponsive	APAP 65 g (7 hr prior)	Before admission: naloxone 1.5 mg IV Admission: intubation, NAC 6.25 mg/kg/hr IV, sodium bicarbonate, fomepizole 15 mg/kg IV 3 hr after admission: hemodialysis × 3 hr, NAC increased to 18.75 mg/kg/hr IV × 3 days	At 7 hr: 883 μg/mL (undetectable on admission day 3)	Admission day 2: AST, 1,791 U/L ALT, 743 U/L (normalized ≈4 days after admission)	Patient was discharged after 4 days of hospitalization without systemic, hepatic, or mental abnormalities
Link et al. [15] (2020)	45	Male	-	-	4 hr after admission: NAC 150 mg/kg IV, then 12.5 mg/kg/hr IV, then 18.75 mg/kg/hr IV at 8 hr (stopped at 35 hr), fomepizole 15 mg/kg IV, then 15 mg/kg IV at 16 hr, then 10 mg/kg IV at 28 hr 8 hr after admission: CRRT until 23 hr	4 hr after admission: 791 μg/mL (< 74.5 μg/mL at 28 hr after admission)	ALT remained WNL	No liver injuries were sustained, and liver enzymes, neurocognition returned to baseline on day 8
Hidalgo et al. [16] (2020)	18	Female	Lethargy	APAP 46 g (10 hr prior)	10 hr: NAC IV After NAC: fomepizole 15 mg/kg IV, hemodialysis × 4 hr After dialysis: fomepizole 10 mg/kg IV After second dose of fomepizole: CRRT	At 10 hr: > 250 μg/mL (undetectable 18 hr after dialysis)	At 58 hr: AST, 3,820 IU/L ALT, 4,346 IU/L (normalized at unknown time)	Discharged after 6 days without issues
Shah and Beuhler [17] (2020)	33	Male	AUD Abdominal pain, tachycardia, tachypnea	APAP 25 g (taken over past 2 days)	9 hr after admission: NAC 150 mg/kg IV, then 12 mg/kg/hr IV × 4 hr, then 6.25 mg/kg/hr IV × 9 hr 11 hr after admission: fomepizole 15 mg/kg IV	Admission: 337 μg/mL (20 μg/mL at 42 hr)	Admission day 3: AST, 198 U/L ALT, 301 U/L	Patient was medically cleared and did well

(Continued on the next page)

Table 1. (Continued)

Study	Age (yr)	Sex	Medical history, presentation, timing	Ingestion dose and timing (if known)	Intervention and timing of interventions after ingestion unless stated otherwise (if known)	Laboratory finding[a]		Clinical course
						APAP level	Liver enzyme level	
Kiernan et al. [18] (2019)	64	Female	Comatose, anion gap metabolic acidosis, lactatemia	APAP 104 g, diphenhydramine 5.2 g (3 hr prior)	3 hr: intubation, epinephrine 100 μg IV, dopamine 10 μg/kg/min IV then 15 μg/kg/min IV, norepinephrine 10 μg/min IV After initial therapy: sodium bicarbonate IV 2 hr after first APAP level: NAC 200 mg/kg IV during dialysis 300 mg/kg IV at 15 hr and 100 mg/kg IV at 16 hr 7 hr after first APAP level: fomepizole 15 mg/kg IV then 10 mg/kg IV at 15 hr 10 hr after first APAP level: hemodialysis ×6 hr	At 3 hr: 1,017 μg/mL (< 15 μg/mL at 48 hr)	At 3 hr: AST, 21 IU/L; ALT, 99 IU/L	Complete resolution of metabolic acidosis and coma without hepatitis or sequelae, discharged 8 days after admission
Villano et al. [19] (2015)	28	Male	Depressed LOC, unresponsiveness, tachycardia, tachypnea, acidemia	Unknown dose (16 hr prior)	Timing unknown: fomepizole (unknown dose), NAC ×45 hr	At 16 hr: 616 μg/mL	At 16 hr: ALT, 27 U/L After initial labs: AST, 67 U/L	Recovery was made without hemodialysis
Zell-Kanter et al. [20] (2013)	59-	Female	Depression Obtunded, agonal respirations	-	Timing unknown: intubation, activated charcoal 1/kg, sodium bicarbonate 1–2 mEq/kg, NAC ×59 hr, naloxone, fomepizole ×1 dose (unknown dose)	Several hours after initial readings: 1,193 mg/L	Admission day 2: AST, 3,150 U/L ALT, 2,780 U/L (normalized on admission day 4)	Patient treated successfully by admission day 4

APAP, acetaminophen; NAC, N-acetylcysteine; IV, intravenous; CRRT, continuous renal replacement therapy; PO, per oral; AST, aspartate transaminase; ALT, alanine transaminase; SUD, substance use disorder; IN, intranasal; AUD, alcohol use disorder; AKI, acute kidney injury; LOC, loss of consciousness; RUQ, right upper quadrant; CKD, chronic kidney disease; WNL, within normal limits.
[a]Highest reported level.

more background on the included case studies. The studies sometimes reported a range of demographic characteristics, including age (mean age, 38.1 years), sex (female sex, 65%), dose of APAP ingested (median dose, 53.1 g), time to presentation from ingestion (median time, 5 hours), and various comorbidities among participants.

Among the 24 case reports and case series, most patients were deemed to have a high risk of hepatotoxicity and presented with massive APAP ingestion without other coingestants. In studies that reported the time between ingestion and presentation, the patient distribution was relatively equal between those who presented before and after the 8- to 10-hour mark after APAP ingestion. The most prevalent treatment regimen in these high-risk patients was dual therapy (NAC and fomepizole), with fomepizole most often initiated as a single dose more than 8 to 10 hours after APAP ingestion. Triple therapy (NAC, fomepizole, and dialysis) was also reported, often without indicating a preference for the timing of fomepizole, but in those cases, multidose regimens were seen more frequently than single-dose regimens. Single-dose-fomepizole monotherapy was also used in four of these cases, all of whom had APAP-aminotransferase multiplication product (APAP×AT) of more than 10,000 and developed shock or multiorgan dysfunction following chronic or repeated supratherapeutic APAP ingestions. The few patients who were at possible risk, low risk, or had an undetermined risk of hepatotoxicity typically presented within 8 to 10 hours after consumption and were successfully managed with either dual or triple therapy. In all but one patient, favorable clinical and laboratory outcomes were reported. Specifically, there was consistent resolution of neurocognitive, acid-base, and hemodynamic abnormalities without significant deterioration, organ failure, or the need for a liver transplant. APAP levels declined steadily throughout each patient's clinical course, and the levels of liver enzymes and markers of liver function rose initially before recovering shortly after fomepizole initiation.

Regarding the ordinal aspect of each patient's therapeutic reg-

Table 2. Summary of included case series about fomepizole use in acetaminophen overdose

Study	Age (yr)	Sex	Medical history and presentation	Ingestion dose and timing (if known)	Timing of interventions after ingestion unless stated otherwise (if known)	Laboratory finding[a]		Clinical course
						APAP level and APAP × AT	Liver enzyme level	
Link et al. [21] (2022) (n = 7)	47	Female	Lightheadedness, confusion, hypoglycemia	APAP 1 g every 4 hr × 2–3 days (12 hr prior)	16 hr: NAC 150 mg/kg IV then 12.5 mg/kg/hr IV Several hours after NAC: intubation, norepinephrine, vasopressin, angiotensin II After intubation/vasopressors: CRRT 24 hr: fomepizole 15 mg/kg IV × 4 doses at 24, 36, 57, 70 hr	16 hr: 10.8 μg/mL APAP × AT, 26,322	29 hr: ALT, 3,325 U/L (118 U/L at 258 hr) 41 hr: AST, 9,305 U/L (77 U/L at 258 hr)	AKI developed during hospitalization and was managed with CRRT Patient developed hypotension requiring vasopressors A full recovery was made, and no liver failure was observed at discharge
	43	Female	HIV, SUD, depression, epilepsy Hemorrhoidal pain	APAP 5 g (12 hr prior)	19 hr: NAC 150 mg/kg IV then 12.5 mg/kg/hr IV 31 hr: fomepizole 15 mg/kg IV	19 hr: 18 μg/mL APAP × AT, 50,620	30 hr: ALT, 3,555 U/L AST, 12,655 U/L 114 hr: ALT, 382 U/L AST, 64 U/L	A full and rapid recovery was made, and the patient was discharged without new impairments Transaminases recovered quickly
	18	Female		APAP 25 g, diphenhydramine 300–600 mg (4 hr prior)	4 hr: NAC 150 mg/kg IV then 12.5 mg/kg IV 8 hr: fomepizole 15 mg/kg IV × 2 doses at 8 and 21 hr	4 hr: 159 μg/mL APAP × AT, 2,560	4 hr: 159 μg/mL	No significant liver injury occurred
	51	Female	ESRD, diabetes, COPD, depression, intimate partner violence	APAP 6.5–9.75 g (4 hr prior)	4 hr: NAC 150 mg/kg IV then 12.5 mg/kg/hr IV 16 hr: fomepizole 15 mg/kg IV then 10 mg/kg IV at 27 hr 20-24 hr: hemodialysis	4 hr: 313 μg/mL APAP × AT, 5,008	18 hr: ALT, 693 U/L AST, 1,450 U/L 93 hr: ALT, 240 U/L AST, 84 U/L	Transaminases improved quickly
	15	Female	Diabetes, history of self-harm Decreased responsiveness,	-	18 hr: NAC 150 mg/kg IV then NAC 12.5 mg/kg/hr IV 26 hr: fomepizole 15 mg/kg IV	18 hr: 89 μg/mL APAP × AT, 1,602	22 hr: ALT, 55 U/L AST, 142 U/L 43 hr: ALT, 36 U/L AST, 37 U/L	No outcomes documented
	16	Female	Abdominal pain, vomiting	APAP 15 tablets (1 day prior; unknown dose)	24 hr: NAC 150 mg/kg IV then NAC 12.5 mg/kg/hr IV 32 hr: fomepizole 15 mg/kg IV After fomepizole: vitamin K 10 mg IV and lactulose	24 hr: 24 μg/mL APAP × AT, 97,488	32 hr: ALT, 3,812 U/L AST, 6,222 220 hr: ALT, 284 U/L AST, 32 U/L	Transaminases continued to rise despite NAC but were managed with fomepizole Transaminases improved, and a full recovery was made
	64	Female	AUD Unresponsive	-	Before admission: intubation 5 hr: NAC 150 mg/kg IV then NAC 12.5 mg/kg/hr IV infusion 13 hr: fomepizole 15 mg/kg IV then 10 mg/kg IV at 23 hr	6 hr: 411.5 μg/mL APAP × AT, 83,268	45 hr: AST, 1,103 U/L 52 hr: ALT 1,260 U/L 81 hr: AST, 160 U/L ALT, 721 U/L	Transaminases started to improve without clinical signs of deterioration by hour 81

(Continued on the next page)

Table 2. (Continued)

Study	Age (yr)	Sex	Medical history and presentation	Ingestion dose and timing (if known)	Timing of interventions after ingestion unless stated otherwise (if known)	Laboratory finding[a]		Clinical course
						APAP level and APAP × AT	Liver enzyme level	
Kaiser et al. [22] (2021) (n = 4)	30	Female	Shock, multiorgan dysfunction, hypotension	Acute APAP ingestion (timing unknown)	Timing unknown: fomepizole 15 mg/kg IV × 1 dose	31 hr: APAP × AT, 31,020	-	Improved after fomepizole infusion and survived to hospital discharge
	46	Male	Shock, multiorgan dysfunction, hypotension	Repeated supratherapeutic APAP ingestion (timing unknown)	Timing unknown: fomepizole 15 mg/kg IV × 1 dose	Initial: APAP × AT, 90,252	-	Improved after fomepizole infusion and survived to hospital discharge
	44	Male	Shock, multiorgan dysfunction, hypotension	Repeated supratherapeutic APAP ingestion (timing unknown)	Timing unknown: fomepizole 15 mg/kg IV × 1 dose	Initial: APAP × AT, 231,763	-	Improved after fomepizole infusion and survived to hospital discharge
	54	Male	Shock, multiorgan dysfunction, hypotension	Chronic APAP ingestion (timing unknown)	Timing unknown: fomepizole 15 mg/kg IV × 1 dose	Initial: APAP × AT, 39,592		Improved after fomepizole infusion and survived to hospital discharge
Shah et al. [23] (2021) (n = 2)	44	Female	SUD Altered mental status	-	Admission: crystalloid 40 mL/kg IV 19 hr after admission: NAC 140 mg/kg PO then 70 mg/kg PO every 4 hr, fomepizole 15 mg/kg IV 3 Days after admission: NAC 150 mg/kg IV then 15 mg/kg/hr IV	13 hr after admission: 108.1 μg/mL	3 Days after admission: ALT, > 5,000 IU/L AST, 8,372 IU/L	Shock developed during admission and required IV NAC Patient survived to discharge
	56	Female	Breast cancer, bipolar disorder, chronic pain syndrome Altered mental status	-	Admission: crystalloid 40 mL/kg IV After crystalloid: NAC 150 mg/kg IV then 15 mg/kg/hr IV 19 hr after initial labs: fomepizole 15 mg/kg IV	19 hr after initial labs: > 300 μg/mL	4 Days after admission: ALT, 4,553 IU/L AST, 2,579 IU/L	Patient survived to discharge
Shah et al. [24] (2020) (n = 3)	62	Female	SIRS, lactic acidosis, arterial acidemia	Unknown dose (5 hr prior)	Timing unknown: NAC, fomepizole (unknown dose)	5 hr: 1,014 μg/mL APAP × AT, 521,202	5 hr: ALT, 10 IU/L AST, 13 IU/L	Survived to hospital discharge without liver transplant
	56	Female	SIRS, lactic acidosis, arterial acidemia	-	Timing unknown: crystalloid 40 mL/kg IV, NAC, fomepizole (unknown dose)	Timing unknown: 298 μg/mL APAP × AT, 97,148	Timing unknown: ALT, 326 IU/L AST, 245 IU/L	Survived to hospital discharge without liver transplant
	58	Female	SIRS, lactic acidosis, arterial acidemia	-	Timing unknown: crystalloid 4L IV, NAC, fomepizole (unknown dose)	Timing unknown: 22.4 μg/mL APAP × AT, 133,750	Timing unknown: transaminases 3,466–11,001 IU/L	Survived to hospital discharge without liver transplant
Rampon et al. [25] (2020) (n = 6)	49	Female	Depression Encephalopathy, salicylates 45 mg/dL	APAP, benzodiazepines, other unknown coingestants (4 hr prior)	4 hr: NAC 150 mg/kg IV then 12.5 mg/kg/hr IV 6 hr: fomepizole 15 mg/kg IV then 10 mg/kg IV every 12 hr 36 hr: hemodialysis × 1 session	4 hr: 140.8 μg/mL (undetectable after hemodialysis)	62 hr: ALT, 79 U/L AST, 51 U/L 110 hr: ALT, 60 U/L AST, 26 U/L	Persistently elevated APAP levels were seen despite NAC and fomepizole; hemodialysis started to address this No biochemical or clinical evidence of liver failure developed Therapy was well tolerated

(Continued on the next page)

Table 2. (Continued)

Study	Age (yr)	Sex	Medical history and presentation	Ingestion dose and timing (if known)	Timing of interventions after ingestion unless stated otherwise (if known)	Laboratory finding[a]		Clinical course
						APAP level and APAP×AT	Liver enzyme level	
	14	Female	Encephalopathy, lethargy	APAP, diphenhydramine (4 hr prior)	4 hr: NAC 150 mg/kg IV then 12.5mg/kg/hr IV 6 hr: fomepizole 15 mg/kg IV then 10 mg/kg IV every 12 hr	6 hr: 251.6 μg/mL	4 hr: AST, 19 U/L 102 hr: ALT, 12 U/L	Persistently elevated APAP levels were seen despite NAC; fomepizole was started to address this No biochemical or clinical evidence of liver failure developed Therapy was well tolerated
	9	Male	-	-	4 hr: NAC 150 mg/kg IV then 12.5 mg/kg/hr IV 6 hr: fomepizole 15 mg/kg IV then 10 mg/kg IV every 12 hr	4 hr: 281.7 μg/mL	20 hr: AST, 47 U/L 40 hr: ALT, 41 U/L	Persistently elevated APAP levels were seen despite NAC; fomepizole was started to address this No evidence of liver failure was observed Therapy was well tolerated
	15	Female	Depression, prior suicide attempts Nausea, vomiting	APAP 50 g, unknown coingestants (2 hr prior)	2 hr: activated charcoal, NAC 150 mg/kg IV then 12.5 mg/kg/hr IV 8.5–9 hr: NAC increased to 18.75 mg/kg/hr IV, fomepizole 15 mg/kg IV then 10 mg/kg IV every 12 hr	8.5 hr: 311.9 μg/mL	16 hr: ALT, 15 U/L AST, 16 U/L	Persistently elevated APAP levels were seen despite NAC; fomepizole was started to address this No evidence of liver failure was observed Therapy was well tolerated
	42	Female	Nausea, lethargy, tachycardia	APAP 100 g, ibuprofen 40 g, loperamide 400 mg (5 hr prior)	5 hr: NAC 150 mg/kg IV then 12.5 mg/kg/hr IV 6 hr: fomepizole 15 mg/kg IV then 10 mg/kg IV every 12 hr	5 hr: 201.8 μg/mL	12 hr: ALT, 29 U/L AST, 54 U/L	Persistently elevated APAP levels were seen despite NAC; fomepizole was started to address this No evidence of liver failure was observed Therapy was well tolerated
	15	Female	Nausea, RUQ pain, abdominal tenderness	APAP 50–62.5 g (2 hr prior)	2 hr: NAC 150 mg/kg IV then 12.5 mg/kg/hr IV ×21 hr 5.5 hr: fomepizole 15 mg/kg IV	4 hr: 361 μg/mL	13 hr: ALT, 19 U/L AST, 10 U/L	No evidence of hepatocellular injury or liver failure was observed Therapy was well tolerated

(Continued on the next page)

Table 2. (Continued)

Study	Age (yr)	Sex	Medical history and presentation	Ingestion dose and timing (if known)	Timing of interventions after ingestion unless stated otherwise (if known)	Laboratory finding[a]		Clinical course
						APAP level and APAP × AT	Liver enzyme level	
Appelqvist and Lindeman [26] (2017) (n = 5)	21	Male	Disorientation, metabolic acidosis	APAP 100 g (2.5 hr prior)	4 hr: activated charcoal 50 g, NAC 12.5 mg/kg/hr IV then 25 mg/kg/hr IV infusion for 7 hr 5.5 hr: fomepizole 15 mg/kg IV 7 hr: CRRT, NAC 25 mg/kg/hr IV for 16 hr then 12.5 mg/kg/hr IV	2.5 hr after admission: 687 (unit unknown) 24.5 hr after admission: 7 (unit unknown)	8.5 hr after admission: ALT, 70 (unit unknown)	Patient recovered without liver transplant
	64	Female	Coma, anion gap metabolic acidosis	APAP 104 g, diphenhydramine 52 g (timing unknown)	Timing unknown: NAC 100 mg/kg IV then 200 mg/kg IV then 300 mg/kg IV 7 hr after first APAP level: fomepizole 15 mg/kg IV then 10 mg/kg IV 8 hr later After fomepizole: hemodialysis	Admission: 1,017 (unit unknown) 41.5 hr after initial tests: < 15 (unit unknown)	Admission: ALT, 99 U/L	Patient recovered without liver transplant
	33	Male	AUD Abdominal pain	APAP 25 g (2 days prior)	9 hr: NAC 150 mg/kg IV then 12.5 mg/kg/hr IV × 4 hr then 6.25 mg/kg/hr IV through 9 hr then 15 mg/kg/hr IV until day 3 11 hr: fomepizole 15 mg/kg IV	Admission: 337 (unit unknown) 42 hr after admission: 20 (unit unknown)	Admission day 2: ALT, 301 AST, 198 (units unknown)	Patient recovered without liver transplant
	55	Male	Diabetes, CKD stage III ECG QRS 114 msec	APAP 65 g (timing unknown)	Timing unknown: NAC 6.25 mg/kg/hr IV then 18.75 mg/kg/hr IV, fomepizole 15 mg/kg IV, hemodialysis × 3 hr, sodium bicarbonate	7 hr after admission: 883 (unit unknown) Admission day 3: undetectable	Admission day 2: ALT, 743 AST, 1,791 (units unknown)	Patient recovered without liver transplant
	18	Female	Lethargy	APAP 46g (10 hr prior)	Timing unknown: NAC PO and IV + hemodialysis for 4 hr followed by hemodialysis 12 hr continuous, fomepizole 15 mg/kg IV and changed to 10 mg/kg IV after dialysis	Admission: > 250 (units unknown)	48 hr after admission: AST, 3,820 ALT, 4,346 (units unknown) Admission day 6: normalized	Patient recovered without liver transplant

APAP, acetaminophen; APAP × AT[b], acetaminophen-aminotransferase multiplication product; NAC, N-acetylcysteine; IV, intravenous; CRRT, continuous renal replacement therapy; ALT, alanine transaminase; AST, aspartate transaminase; AKI, acute kidney injury; SUD, substance use disorder; ESRD, end-stage renal disease; COPD, chronic obstructive pulmonary disease; AUD, alcohol use disorder; PO, per oral; SIRS, systemic inflammatory response syndrome; RUQ, right upper quadrant; CKD, chronic kidney disease; ECG, electrocardiogram.

[a]Hours after ingestion if known unless stated otherwise. [b]When the APAP nomogram cannot be used (i.e., when the time after ingestion is unknown), an APAP × AT multiplication product was calculated by multiplying APAP and ALT or AST levels (whichever is higher). A value above the threshold of 10,000 mg/L × U/L suggests a potential risk of hepatotoxicity and the need to initiate therapy for APAP toxicity.

imen, fomepizole was most commonly started either at the same time as NAC or after NAC but before dialysis was initiated. On numerous occasions, the use of fomepizole was rationalized by either persistently elevated or increasing APAP and liver enzyme levels or the severity of patient presentation and possible coingestants prompting concerns about a significant risk of morbidity or near-term mortality. In all cases, fomepizole was most frequently administered at the dose standard for treating toxic alcohol poisoning (15 mg/kg intravenous [IV], followed by 10 mg/kg IV every 12 hours as needed). No adverse events associated with fomepizole were reported in any of these case reports or case series.

Our search also yielded a single double-blinded, randomized, controlled crossover trial, conducted by Kang et al. [27], which examined the effect of fomepizole alone on APAP levels in overdoses. The study involved five healthy volunteers. Participants received APAP alone or with fomepizole, with a 2-week washout period between treatments. Fomepizole significantly reduced the formation of oxidative metabolites, indicating its potential to mitigate APAP toxicity. However, that study's short timeline did not capture effects during peak hepatotoxicity, which occurs 72 to 96 hours after ingestion. No adverse events were reported.

A secondary analysis of the ToxIC (Toxicology Investigators Consortium) database by Filip et al. [28] between January 2015 and July 2020 revealed that when fomepizole was explicitly used to manage APAP-overdose patients (n = 25) with markers of criti-

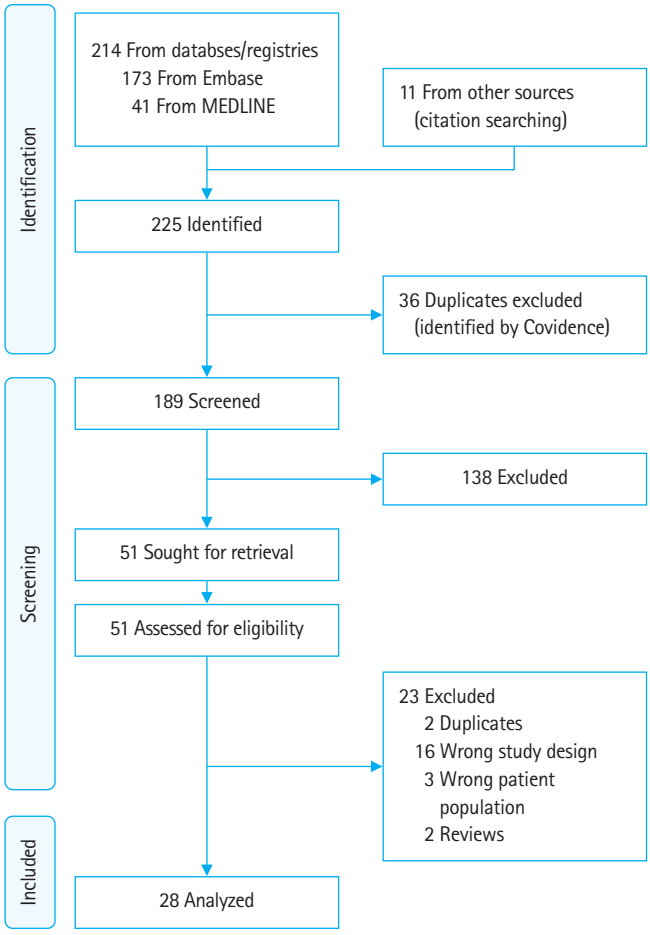


Fig. 1. PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flowchart for the literature search in MEDLINE and Embase.

cal illness (hepatotoxicity/coagulopathy, metabolic acidosis, need for intubation/vasopressors/continuous renal replacement therapy), an overall mortality rate of 24% (6 of 25) was observed. It is interesting to note that this is higher than the mortality rates observed in the case reports and case series included in our review. However, the patients included in that study exhibited severely toxic clinical presentations, so its mortality rate was confounded by critical illness. Filip et al. [28] also explored prescriber attitudes about the use of fomepizole in that context. Most prescribers indicated that they would consider the use of fomepizole at the standard dose for managing toxic alcohol poisoning in conjunction with dialysis in both acute and chronic APAP overdoses. Most prescribers also agreed that fomepizole is safe in patients with significant metabolic acidosis or hepatotoxicity. No additional patient outcomes were derived from that study.

Two abstracts from conference proceedings are also included in this review. An institutional review of the use of fomepizole and NAC in patients with massive APAP ingestion and without co-ingestion of hepatotoxic substances (n = 41) by Stott et al. [29] demonstrated that patients who started fomepizole and NAC on the same day had significantly lower peak AST levels and total days on NAC therapy than those who started the treatments on different days. A trend toward fewer deaths and major outcome designations was also observed in the same-day group. Vottero and Mégarbane [30] performed an observational retrospective cohort study in which mortality was observed in one of six patients who presented to a non-liver intensive care unit with acute liver failure and received both fomepizole and NAC.

DISCUSSION

To our knowledge, this review includes all the latest findings since July 2021 and adds to the available body of evidence summarized by recent systematic and scoping reviews by Mullins et al. [31] and Pourbagher-Shahri et al. [2]. It presents evolving insights into the hepatoprotective mechanisms, optimal dosing strategies, and safety profile of fomepizole in treating APAP poisoning. Prior to this narrative review, the benefits of adding fomepizole to NAC had been suggested by various *in vitro* and *in vivo* studies, but human data were quite limited. Thus, it remained unclear in both systematic or scoping reviews whether fomepizole alone or in combination with NAC was a primary contributor to each patient's positive clinical outcomes [2,31]. Including studies from previous scoping reviews and new cases allowed for a more comprehensive overview of fomepizole use in APAP poisoning, revealing clinical trends and indications that were not addressed in earlier reviews.

Case series and case reports continue to offer insights into how fomepizole fits into patient care in the setting of APAP overdose in a real-life scenario, but observational cohort studies are needed to better understand the clinical outcomes of fomepizole in APAP toxicity management. Most care teams in these case reports did not pre-emptively initiate fomepizole with NAC upon initial toxic presentation. Fomepizole was typically started when APAP concentrations, liver enzyme markers, or a patient's clinical status were not amenable to an IV NAC infusion. Instances in which fomepizole was initiated close to or at the same time as NAC seemed to occur often in patients who eventually required dialysis. This could imply that the severity of a patient's toxic presentation factors into whether a clinician chooses to start fomepizole earlier as opposed to later. Beneficial clinical and laboratory effects were observed in these studies, but it remains unclear whether these outcomes can be attributed specifically to fomepi-

zole alone or its use alongside NAC and hemodialysis or if fomepizole offered any significant hepatoprotection. Specifically, APAP concentrations and liver enzyme levels, along with other relevant laboratory markers (lactate, venous blood gas, bilirubin, international normalized ratio), seemed to decrease and become responsive upon the addition of fomepizole, but causality cannot be confirmed. Despite most cases resulting in the resolution of the patient's general clinical presentation, liver/acid-base disturbances, and altered mental status, these studies did not provide in-depth analyses of their clinically meaningful outcomes. Nonetheless, these results suggest fomepizole's potential role in improving hepatotoxicity in ways that reduce near-term mortality and the need for liver transplantation.

It has become clear that fomepizole is being used more liberally by prescribers for the management of APAP-related poisoning. The systematic review by Mullins et al. [31] recommended fomepizole as adjunctive therapy in patients with at least one of the following characteristics of severe poisoning: APAP concentration above the 300 line of the APAP treatment nomogram, metabolic acidosis, positive King's College criteria, or late presentation with evident liver injury and measurable APAP levels. Among the 45 patients in the included case studies, the compiled trends and newly discovered themes across cases provide insight into additional criteria that could be used to determine when fomepizole therapy is warranted and how it could be administered. The severity of a patient's clinical presentation and risk of mortality/morbidity were among the reasons most commonly cited in these case reports for initiating fomepizole as an adjunct to NAC or dialysis. APAP concentrations and liver enzymes or acid-base abnormalities that were not amenable to NAC infusions were also among the reasons fomepizole was started or multiple doses were administered. Although select cases did report that fomepizole monotherapy provided favorable patient outcomes, the limited clinical experience with this regimen precludes fomepizole from replacing NAC as the standard of care. Limited documentation in each case report makes it difficult to identify a specific timeframe in which the drug should be started. However, it is known that metabolism impairment occurs during an APAP overdose that results in a prolonged half-life of 4 to 8 hours [32]. That supports early initiation of fomepizole to take advantage of its hepatoprotective mechanisms during the APAP metabolism and post-metabolism phases, sourced from its CYP2E1 and JNK inhibition, respectively [33]. This consideration was also offered by Stott et al. [29], who showed that starting fomepizole preemptively on the same day as NAC could limit the severity of hepatotoxicity and reduce the time of NAC therapy. Thus, in addition to the recommendations made by Mullins et al. [31], fomepizole should be administered within 4 to 8 hours at the standard dose used in alcohol poisoning (15 mg/kg IV, followed by 10 mg/kg IV every 12 hours as needed) in patients at high risk of immediate mortality/morbidity. This includes those with severe clinical presentations, APAP levels above the Rumack-Matthew nomogram line, APAP × AT > 10,000, and other markers of severe hepatotoxicity (e.g., refractory metabolic acidosis, coagulopathy, encephalopathy, acute kidney injury). Patients with APAP or liver enzyme levels that are not amenable to NAC infusion with or without dialysis should be considered for fomepizole therapy as well. Although fomepizole is removed by dialysis, there are major discrepancies on its dosing in patients receiving dialysis. Up to 42% of survey respondents in the study by Filip et al. [28] suggested that fomepizole redosing is required following dialysis. Fomepizole's drug product monograph suggests that for alcohol intoxication, doses should be administered every 4 hours during dialysis [34].

No adverse events associated with fomepizole were documented in any case reports or case series included in this review. Furthermore, the lack of adverse events observed in the randomized-controlled trial by Kang et al. [27] and the extensive positive historical experience using fomepizole to treat toxic alcohol poisoning support its safe use in patients with APAP poisoning. Some common adverse reactions experienced with fomepizole, according to its drug monograph, include headache, nausea, dizziness, drowsiness, and metallic taste [34]. Less common side effects include abdominal pain, fever, injection site reactions, phlebitis, hypotension, vomiting, diarrhea, and facial flush [34]. A double-blind randomized-controlled trial by Jacobsen et al. [35] showed that in healthy volunteers, single doses of 10 and 20 mg/kg of fomepizole did not elicit any adverse effects. Doses above that caused mild to moderate and short-lived nausea, dizziness, and vertigo. No adverse effects linked to changes in objective readings (pulse, blood pressure, temperature, and blood/urine chemistries) were reported. However, the results of that study are limited by a lack of blinding to higher doses of fomepizole due to its poor palatability.

The cost of fomepizole as a novel therapy in APAP poisoning warrants consideration. Economic analyses have shown the inflation-adjusted cost of a single dose of fomepizole (15 mg/kg) to be US $748 to $1,123, which is similar to a 20-hour infusion of NAC (US $643 to $735) [36–40]. With the cost of liver transplants in the United States and other countries within the Organization for Economic Co-operation and Development (OECD) often exceeding US $150,000 to $200,000 and the cost of contin-

uous renal replacement therapy, based on data from two tertiary intensive care units in Canada, ranging from CAD $3,486 to $5,117 per week, the potential hepatoprotective benefits of fomepizole could justify a more liberal use of this medication [41,42].

Limitations

Despite a promising place for fomepizole in patients with APAP overdoses, these results are confounded by a lack of standardization and consistency in treatment protocols, which hinders our ability to identify the main contributor to each patient's clinical outcome. The assessment of each patient case was also complicated by a lack of documentation of pertinent clinical and laboratory information, which hindered our ability to make general inferences about hepatotoxicity risk stratification, the severity of APAP ingestion and toxic manifestation, the safety of fomepizole therapy, and trends in therapeutic interventions. The small sample sizes and lack of comparators in these case studies also limit their generalizability and undermine our ability to determine causative relationships. Selection bias and publication bias are also possible, with unusual or dramatic patient overdose cases being more likely to be published, so these cases probably do not reflect the entire spectrum of APAP overdose events. The included randomized-controlled trial, observational study, and reviews are also limited by small sample sizes and a lack of analyses for additional patient-centered outcomes. It was difficult to appraise the validity and integrity of each study in the abstracts sourced from conference proceedings. Future observational cohort studies would be an appropriate next step and will be essential to further define the hepatoprotective role of fomepizole and its effects on meaningful clinical outcomes in APAP overdose cases.

Conclusions

Human case studies and reports highlight the potential role of fomepizole as an adjunctive therapy in APAP poisoning, particularly in patients for whom intravenous NAC alone is insufficient. Although fomepizole is not a replacement for NAC as the standard of care, it might have contributed to the beneficial outcomes seen in severe poisoning cases, including those with high APAP concentrations, metabolic acidosis, or signs of liver injury. Historical experience with fomepizole in toxic alcohol poisoning and the limited evidence sourced from the randomized, controlled crossover trial suggest that fomepizole exhibits a good safety profile with limited adverse effects. Overall, the evidence supports fomepizole's consideration in select scenarios, especially in cases with a high risk of hepatotoxicity, severely toxic or refractory clinical status, or late presentation after APAP ingestion. However, in cases of rapid clinical deterioration, a low threshold for hemodialysis remains essential, emphasizing the need for further research to support fomepizole's role in APAP toxicity management.

ARTICLE INFORMATION

Author contributions

Conceptualization: AL, SZQL, HC; Data curation: AL, HC, FC; Writing–original draft: FC, AL; Writing–review & editing: all authors. All authors read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

Funding

The authors received no financial support for this study.

Data availability

Data analyzed in this study are available within the article and its supplementary materials.

Supplementary materials

Suppl. 1. Search strategy used for literature search in selected MEDLINE and Embase databases

Suppl. 2. Critical appraisal of included case reports

Suppl. 3. Critical appraisal of included case series

Supplementary materials are available from https://doi.org/10.15441/ceem.25.059.

REFERENCES

1. Health Canada. Acetaminophen [Internet]. Government of Canada; [cited 2024 Dec 17]. Available from: https://www.canada.ca/en/health-canada/services/drugs-medical-devices/acetaminophen.html
2. Pourbagher-Shahri AM, Schimmel J, Shirazi FM, Nakhaee S, Mehrpour O. Use of fomepizole (4-methylpyrazole) for acetaminophen poisoning: a scoping review. Toxicol Lett 2022;355:47–61.
3. Williams EE, Quach D, Daigh A. Massive acetaminophen ingestion managed successfully with N-acetylcysteine, fomepizole, and renal replacement therapy. Clin Nephrol Case Stud 2024;12:22–5.
4. Chary M, Vizzi C, Jimenez X. Delayed antidotal fomepizole in

massive acetaminophen overdose: a case report. Toxicol Commun 2023;7:2250968.
5. Lobo GL, Brenner MA. Adjunctive 4-methylpyrazole therapy in massive acetaminophen poisoning aided in preventing severe liver injury. 2023 American College of Medical Toxicology (ACMT) Annual Scientific Meeting; 2023 Mar 31–Apr 2; San Diego, CA, USA. J Med Toxicol 2023;19:40.
6. Pepin L, Matsler N, Fontes A, Heard K, Flaherty BF, Monte AA. Fomepizole therapy for acetaminophen-induced liver failure in an infant. Pediatrics 2023;152:e2022061033.
7. Ho H, Brambl W. Multiple drugs: various toxicities following paracetamol overdose, and lack of efficacy: case report. React Wkly 2023;1958:395.
8. Kusnik A, Mostafa MR, Hunter N, Sharma R, Jamal H, Hameed M. Fomepizole as an emerging adjunct in treating severe acetaminophen toxicity: a case report and a brief review. J Gastrointestin Liver Dis 2022;31:481–2.
9. Bassi M, Sweat C, Armenian P. Use of endoscopic gastric decontamination and fomepizole in the treatment of massive extended-release acetaminophen overdose complicated by bezoar and persistently elevated APAP level. North American Congress of Clinical Toxicology (NACCT) 2022; 2022 Sep 14–18; San Francisco, CA, USA. Clin Toxicol (Phila) 2022;60(Suppl 2):125.
10. Dion C, Spyres M. Massive acetaminophen ingestion resulting in altered mental status and intubation without metabolic acidosis after early treatment with fomepizole and N-acetylcysteine. North American Congress of Clinical Toxicology (NACCT) 2022; 2022 Sep 14–18; San Francisco, CA, USA. Clin Toxicol (Phila) 2022;60(Suppl 2):85.
11. Chiu MH, Jaworska N, Li NL, Yarema M. Massive acetaminophen overdose treated successfully with N-acetylcysteine, fomepizole, and hemodialysis. Case Rep Crit Care 2021;2021: 6695967.
12. Homan T, Wichman B. Combination therapy with N-acetylcysteine and 4-methylpyrazole for acetaminophen overdose. Chest 2021;160(4 Suppl): A914.
13. Cuninghame S, Lotfy K, Cameron P. Massive acetaminophen overdose with metabolic acidosis refractory to N-acetylcysteine, fomepizole, and renal replacement therapy. Toxicol Rep 2021;8:804–7.
14. Woolum JA, Hays WB, Patel KH. Use of fomepizole, n-acetylcysteine, and hemodialysis for massive acetaminophen overdose. Am J Emerg Med 2020;38:692.
15. Link S, Scalzo A, Eckerle P, et al. Novel use of fomepizole (4-MP) for CYP2E1 inhibition in acetaminophen overdose. North American Congress of Clinical Toxicology (NACCT) 2020; 2020 Sep 10–14. Clin Toxicol (Phila) 2020;58:1170–1.
16. Hidalgo DC, Shah S, Rech M. 949: Use of fomepizole and hemodialysis for massive acetaminophen toxicity: a case report. Crit Care Med 2020;48:454.
17. Shah KR, Beuhler MC. Fomepizole as an adjunctive treatment in severe acetaminophen toxicity. Am J Emerg Med 2020; 38:410.
18. Kiernan EA, Fritzges JA, Henry KA, Katz KD. A case report of massive acetaminophen poisoning treated with a novel "triple therapy": N-acetylcysteine, 4-methylpyrazole, and hemodialysis. Case Rep Emerg Med 2019;2019:9301432.
19. Villano JH, O'Connell CW, Ly BT, Schneir A. Case files from the University of California San Diego health system fellowship coma and severe acidosis: remember to consider acetaminophen. J Med Toxicol 2015;11:368–76.
20. Zell-Kanter M, Coleman P, Whiteley PM, Leikin JB. A gargantuan acetaminophen level in an acidemic patient treated solely with intravenous N-acetylcysteine. Am J Ther 2013; 20:104–6.
21. Link SL, Rampon G, Osmon S, Scalzo AJ, Rumack BH. Fomepizole as an adjunct in acetylcysteine treated acetaminophen overdose patients: a case series. Clin Toxicol (Phila) 2022; 60:472–7.
22. Kaiser SK, Dart RC, Hoyte C. Fomepizole as adjunctive therapy for delayed acute acetaminophen overdose and repeated supratherapeutic ingestion. 2021 American College of Medical Toxicology (ACMT) Annual Scientific Meeting; 2021 Apr 14-16. J Med Toxicol 2021;17:132.
23. Shah KR, Fox C, Geib AJ, et al. Fomepizole as an adjunctive treatment in severe acetaminophen ingestions: a case series. Clin Toxicol (Phila) 2021;59:71–2.
24. Shah KR, Fox C, Geib AJ, et al. 075. Fomepizole as an adjunctive treatment in severe acetaminophen ingestions. 2020 American College of Medical Toxicology (ACMT) Annual Scientific Meeting; 2020 Mar 12–15; New York, NY, USA. J Med Toxicol 2020;16:116–68.
25. Rampon G, Wartman H, Osmon S, Scalzo A. Use of fomepizole as an adjunct in the treatment of acetaminophen overdose: a case series. Toxicol Commun 2020;4:1–4.
26. Appelqvist Å, Lindeman E. Fomepizole, dialysis and an increased dose of N-acetylcysteine in a case of massive paracetamol ingestion. 37th International Congress of the European Association of Poisons Centres and Clinical Toxicologists (EAPCCT); 2017 May 16-19; Basel, Switzerland. Clin Toxicol (Phila) 2017;55:461.

27. Kang AM, Padilla-Jones A, Fisher ES, et al. The effect of 4-methylpyrazole on oxidative metabolism of acetaminophen in human volunteers. J Med Toxicol 2020;16:169–76.
28. Filip AB, Berg SE, Mullins ME, Schwarz ES; Toxicology Investigators Consortium (ToxIC). Fomepizole as an adjunctive therapy for acetaminophen poisoning: cases reported to the toxicology investigators consortium (ToxIC) database 2015-2020. Clin Toxicol (Phila) 2022;60:1006–11.
29. Stott M, Cowdery C, Simmons R, et al. Clinical impact of fomepizole as an adjunct therapy in massive acetaminophen overdose. North American Congress of Clinical Toxicology (NACCT) 2023; 2023 Sep 27-Oct 1; Montreal, Canada. Clin Toxicol (Phila) 2023;61(Suppl 2):61.
30. Vottero A, Mégarbane B. Outcome of acetaminophen-poisoned patients admitted to a non-liver intensive care unit with liver failure: an observational cohort study. 43rd International Congress of the European Association of Poisons Centres and Clinical Toxicologists (EAPCCT) 2023; 2023 May 23–26; Palma de Mallorca, Spain. Clin Toxicol (Phila) 2023; 61(Suppl 1):116.
31. Mullins ME, Yeager LH, Freeman WE. Metabolic and mitochondrial treatments for severe paracetamol poisoning: a systematic review. Clin Toxicol (Phila) 2020;58:1284–96.
32. Mazaleuskaya LL, Sangkuhl K, Thorn CF, FitzGerald GA, Altman RB, Klein TE. PharmGKB summary: pathways of acetaminophen metabolism at the therapeutic versus toxic doses. Pharmacogenet Genomics 2015;25:416–26.
33. Akakpo JY, Ramachandran A, Duan L, et al. Delayed treatment with 4-methylpyrazole protects against acetaminophen hepatotoxicity in mice by inhibition of c-Jun n-terminal kinase. Toxicol Sci 2019;170:57–68.
34. Fomepizole for injection 1.5 g/1.5 mL (1 g/mL) [Product monograph]. SteriMax Inc; 2018.
35. Jacobsen D, Sebastian CS, Blomstrand R, McMartin KE. 4-Methylpyrazole: a controlled study of safety in healthy human subjects after single, ascending doses. Alcohol Clin Exp Res 1988;12:516–22.
36. Brent J. Fomepizole for ethylene glycol and methanol poisoning. N Engl J Med 2009;360:2216–23.
37. Darracq MA, Rentmeester LL, Clark RF, Tomaszewski CA, Schneir AB, Cantrell FL. Cost of hemodialysis versus fomepizole-only for treatment of ethylene glycol intoxication. Clin Toxicol (Phila) 2013;51:188.
38. Anseeuw K, Sabbe MB, Legrand A. Methanol poisoning: the duality between 'fast and cheap' and 'slow and expensive'. Eur J Emerg Med 2008;15:107–9.
39. Lavonas EJ, Beuhler MC, Ford MD, et al. Intravenous administration of N-acetylcysteine: oral and parenteral formulations are both acceptable. Ann Emerg Med 2005;45:223–4.
40. Marchetti A, Rossiter R. Managing acute acetaminophen poisoning with oral versus intravenous N-acetylcysteine: a provider-perspective cost analysis. J Med Econ 2009;12:384–91.
41. van der Hilst CS, Ijtsma AJ, Slooff MJ, Tenvergert EM. Cost of liver transplantation: a systematic review and meta-analysis comparing the United States with other OECD countries. Med Care Res Rev 2009;66:3–22.
42. Srisawat N, Lawsin L, Uchino S, Bellomo R, Kellum JA; BEST Kidney Investigators. Cost of acute renal replacement therapy in the intensive care unit: results from The Beginning and Ending Supportive Therapy for the Kidney (BEST Kidney) study. Crit Care 2010;14:R46.

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):28-36
https://doi.org/10.15441/ceem.24.289

Review Article

Public awareness of medical emergency telephone numbers: a scoping review

Alexei A. Birkun

Department of General Surgery, Anaesthesiology, Resuscitation and Emergency Medicine, Medical Institute named after S.I. Georgievsky of V.I. Vernadsky Crimean Federal University, Simferopol, Russia

pISSN: 3022-1811
eISSN: 2383-4625

Prompt activation of emergency medical services (EMS) constitutes the fundamental component of bystander response to time-dependent health crises. A clear understanding of the public ability to access EMS can help to guide interventions aimed at enhancing community preparedness for emergencies. This review was conducted to summarize studies that examined public knowledge of emergency phone numbers. This scoping review encompassed articles published since 2004 that reported the proportion of subjects who knew emergency phone numbers. Data sources included PubMed, Google Scholar, and references of included articles. Relevant data from eligible publications were extracted manually to an author-developed data-charting sheet and analyzed descriptively. Forty-eight articles were analyzed. Reported studies, mostly cross-sectional surveys, were conducted in 26 countries, including 16 high-income, 9 middle-income, and 1 low-income country. The percentage of subjects who knew emergency numbers varied from 0% to 97.8% (median, 64.3%; interquartile range [IQR], 32.8%–80.0%). For high-income countries, the median was significantly higher than for low- and middle-income nations (69.6% [IQR, 54.1%–84.2%] vs. 34.6% [IQR, 19.4%–61.5%], P=0.003). The studies were generally inconsistent regarding the association of subjects' sociodemographic factors with knowledge of emergency numbers, suggesting the existence of geography-specific patterns. Available studies observed low community knowledge of emergency numbers, especially in low- and middle-income countries, and suggest that the problem has a global scale. Further research efforts are required to determine the best strategies for enhancing the public ability to access EMS.

Keywords Awareness; Emergencies; Emergency medical services; Knowledge; Telecommunications

Received: July 19, 2024
Revised: October 30, 2024
Accepted: December 12, 2024

Correspondence to:
Alexei A. Birkun
(birkunalexei@gmail.com)

Capsule Summary

What is already known

Studies on public awareness of emergency phone numbers are critical as they uncover and measure the problem of community unpreparedness to respond to acute medical events and may guide interventions to enhance public awareness of how to activate emergency services rapidly. However, seemingly no attempts have been made to summarize and systematically map related evidence.

What is new in the current study

Available studies confirm the alarmingly low community knowledge of emergency phone numbers and imply that the problem has a global scale. Approximately 3 out of 10 people in high-income countries and 6 out of 10 people in the low- and middle-income countries would be unable to call emergency medical services because they do not know what number to dial. The non-absolute public knowledge of how to activate emergency medical services constitutes a risk factor for potentially preventable deaths and disabilities.

How to cite this article:

Birkun AA. Public awareness of medical emergency telephone numbers: a scoping review.
Clin Exp Emerg Med 2026;13(1):28-36
https://doi.org/10.15441/ceem.24.289

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

© 2026 The Korean Society of Emergency Medicine

INTRODUCTION

Activation of emergency medical services (EMS) by dialing the local emergency phone number immediately upon recognition of a problem constitutes an integral component of bystander response in a health crisis [1,2]. Trained dispatchers or telecommunicators may guide the lay rescuer to check the victim's condition, promptly dispatch professional help, and, depending on the system's organization, may also give instructions on how to provide help before the arrival of professionals at the scene. Rapid and direct calls to EMS are associated with shorter time to advanced medical interventions and improved outcomes in time-dependent medical events, including stroke, myocardial infarction, and cardiac arrest [3–5]. Conversely, when bystanders are unable to call EMS because they do not know what number to dial, it can have detrimental effects on the chain of survival.

A number of studies have examined public awareness of emergency phone numbers, demonstrating considerable variation in knowledge depending on geography and sociodemographic characteristics [6,7]. Such studies are important, as they uncover and measure the problem of community unpreparedness to respond to acute medical events and may guide interventions to enhance public awareness of how to rapidly activate emergency services.

However, it seems that no attempts have been made to date to summarize and systematically map related evidence on a global scale. This study aimed to provide a scoping review of original studies that reported public knowledge of emergency phone numbers and factors associated with this knowledge.

METHODS

Ethics statement

This review was carried out and reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines (Suppl. 1) [8]. The review protocol was not preregistered or published. All data that support the findings of this research are openly available from the Mendeley Data repository [9].

Study selection

The review included peer-reviewed English-language articles published since 2004 that reported the number and/or percentage of subjects who knew the phone number(s) for activating EMS. Both cross-sectional observational studies (questionnaire surveys) and interventional studies (evaluating the effects of educational interventions) were considered for inclusion.

The search was performed in May 2024 in PubMed and Google Scholar. The detailed search strategy for PubMed is provided in Suppl. 2. Search results were screened based on the articles' titles and abstracts, and relevant records that appeared to meet the inclusion criteria were exported to Zotero ver. 6.0.36 (Corporation for Digital Scholarship) reference management program. After removing duplicates, full texts were retrieved and evaluated for eligibility. Additionally, relevant papers were searched in reference lists of all included articles and through the Google search engine. Studies that did not contain the key quantitative data, as well as overlapping papers with duplicate results were excluded.

For eligible studies, the following information was extracted and transferred manually to a data-charting sheet [9]: author(s), publication year, study design, study geography, study participants, sample size, data collection settings, type of a question to evaluate the knowledge of emergency numbers, percentage of subjects who knew emergency numbers, factors found to be associated and not associated with knowledge of emergency numbers, and recommendations formulated by the authors based on the research results. In cases when a publication did not report a percentage of subjects who knew emergency numbers (full percentage with one decimal), a relative value was calculated based on available absolute data. For studies that reported only the percentage of those who did not know emergency numbers, the rest of the studied sample was presumed to know the number. For interventional studies that reported knowledge of emergency numbers before and after an intervention, only baseline measures were considered representative of actual knowledge in a population.

Descriptive analysis was applied to summarize and report the findings. The methodological quality of the included studies was not assessed in this review.

RESULTS

Fig. 1 shows the flow diagram of the study selection process. The initial search in PubMed and Google Scholar returned 844 results. Of these, 54 records were considered potentially eligible. After removal of duplicates and the addition of 10 studies found through the hand search of reference lists and web search, 53 full-text articles were assessed for eligibility. Five papers were excluded as noneligible, and eventually, 48 articles were included in the final review [10–57].

The selected articles were published between 2007 and 2024. Respective studies were carried out in 26 countries (12.7% of the 205 world's sovereign states and 13.5% of the 193 member

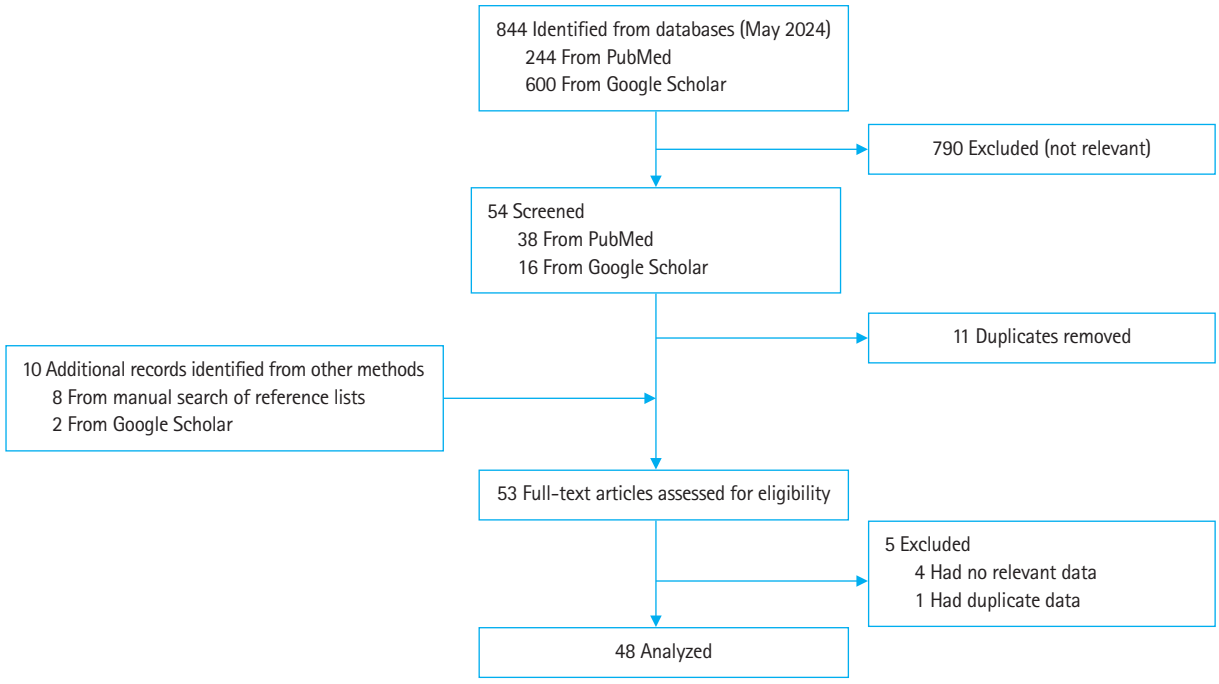


Fig. 1. Study search and selection flow diagram.

states of the United Nations) (Fig. 2), including 16 high-income countries (30 out of 48 studies, 62.5%) [10,14–21,26–30,32–37,39,40,43–45,47,48,52,56,57], 7 upper-middle-income countries (12 studies, 25.0%) [11–13,22–25,41,50,53–55], 2 lower-middle-income countries (2 studies, 4.2%) [31,38], and 1 low-income country (4 studies, 8.3%) [42,46,49,51], according to the World Bank's classification [58]. The number of studies per country varied from one to five.

The majority of publications (39 studies, 81.3%) presented the results of cross-sectional surveys [11–13,15,17–27,29–31,33,35,36,38–54,57], while five studies (10.4%) reported data from before-and-after studies [16,32,34,37,55], two (4.2%) from non-randomized controlled studies, and two (4.2%) from randomized controlled studies [14,56]. In 46 studies (95.8%), the subjects' knowledge of emergency numbers was assessed using self-administered questionnaires or interviews [10–13,15–28,30–57]; in one study (2.1%) the knowledge was evaluated in a simulation scenario of a health emergency [14], and in another (2.1%) both a questionnaire and a simulation scenario were used (the questionnaire survey results were preferred for the data synthesis of the current review) [29]. To assess knowledge of emergency numbers, 24 studies (50.0%) employed open-ended questions [12–14,17,20–24,28,33,34,37,39,43–46,48–50,53–55], 2 (4.2%) used multiple-choice questions [10,40], and 1 (2.1%) used closed-ended question [47], whereas 21 studies (43.8%) did not report the type of inquiry [11,15,16,18,19,25–27,29–32,35,36,38,41,42,51,52,56,57].

Studied populations included the general public (22 studies, 45.8%; mostly adult and usually without exclusion of people with a medical background) [11–13,15,17,20–22,26–28,31,35,38–40,43,45–47,49,52], patients, caregivers, and hospital visitors (10 studies, 20.8%) [16,18,23,24,29,33,41,42,44,54], kindergarten children and school students (9 studies, 18.8%) [10,14,19,30,34,37,50,55,56], and occasionally university students [25], medical students [45], female students of health science colleges [32], primary care physicians of community clinics [36], conscripts [48], marathon runners [53], persons holding a driving license [57], and women who gave birth within the previous 12 months at public health institutions [51] (1 study each, 2.1%). One study provided the percentage of subjects who knew emergency numbers separately for the general public and medical students [45], and hence the total number of data points reported below for the 48 reviewed studies is 49.

Sample size varied from 27 to 9,761 subjects (median, 468; interquartile range [IQR], 151.5–870.5), and the aggregate number of subjects across the reviewed studies was 40,151. The percentage of subjects who knew emergency phone numbers ranged from 0% to 97.8% with a median of 64.3% (IQR, 32.8%–80.0%). This value differed considerably between high-income countries (median, 69.6%; IQR, 54.1%–84.2%) and low- and middle-in-

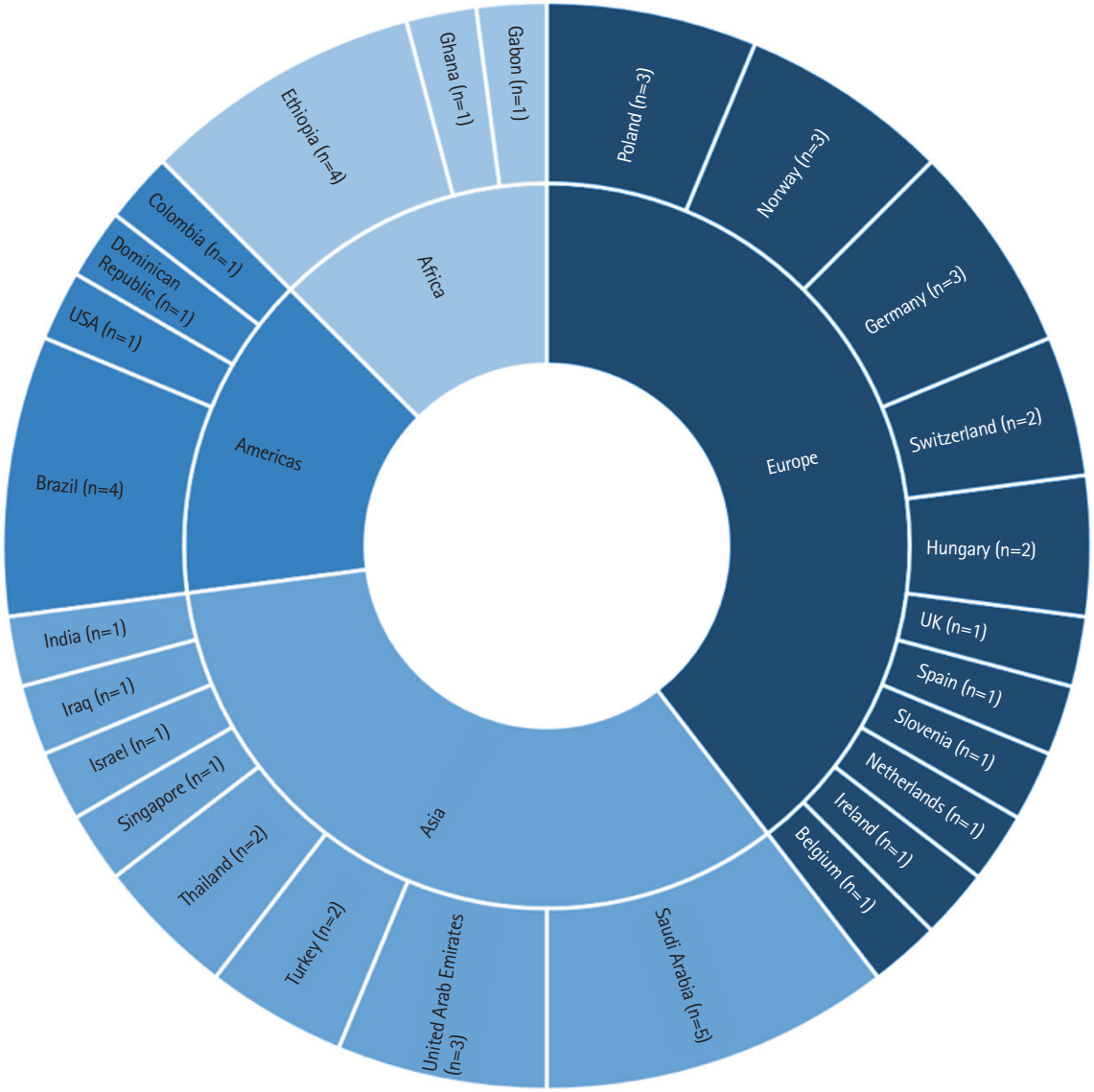


Fig. 2. Distribution of the reviewed studies by global region and country.

come countries (median, 34.6%; IQR, 19.4%–61.5%; P=0.003).

Twenty-two studies (45.8%) investigated the relation of various factors with the subjects' knowledge of emergency phone numbers [11,13,15,19,20,24,27,28,30,34,35,38,39,41,43,44,46,49–53]. Table 1 provides a summary of factors found to be significantly associated or not associated with knowledge of emergency numbers [11,13,15,19,20,24,27,28,30,34,35,38,39,41,43,44,46,49–53].

Nine studies (18.8%), alongside reporting baseline knowledge of emergency numbers, examined the effects of different interventions aimed at improving participants' knowledge and emergency response skills [10,14,16,28,32,34,37,55,56]. The proposed educational interventions, including instructor-led first aid and/or basic life support training for preschool and school children [14,34,37,56], implementation of general healthcare practice-improvement plans based on information services [16], a single physician's advice for adults about correct emergency phone numbers [28], distribution of first aid awareness materials via email among students of health science colleges [32], and use of an educational mobile application for children [55], led to a significant increase in knowledge of EMS phone numbers. In the only interventional study that did not reveal a significant improvement in knowledge, the baseline (pre-intervention) percentage of subjects who knew the emergency number was as high as 97.8% [10].

The majority of articles (34 studies, 70.8%) contained authors'

Table 1. Relationship of subject characteristics with knowledge of EMS phone numbers

Factors associated with knowledge of EMS numbers (no. of studies)	
Yes	No
Age	Age (n = 8) [13,15,20,28,38,39,44,52]
Younger (n = 3) [35,41,43]	
Older (n = 2) [34,50]	
Gender	Gender (n = 7) [13,15,20,28,34,38,52]
Male (n = 1) [43]	
Female (n = 3) [19,39,44]	
Higher educational level (n = 1) [20]	Education level (n = 8) [11,13,28,38,39,43,44,52]
Geographic location of the subjects (a region of a country) (n = 1) [15]	Area of residence (urban vs. rural) (n = 1) [39]
Absence of migration background (n = 1) [52]	Social class (n = 1) [15]
Nationality (Saudi) (n = 1) [43]	Health coverage (public vs. private) (n = 1) [13]
Health education (n = 1) [39]	History of heart disease (n = 1) [44]
Healthcare occupation (n = 1) [38]	History of coronary artery disease (n = 1) [20]
Previous first aid/CPR training (n = 4) [30,34,39,53]	History of myocardial infarction (n = 2) [20,44]
Prior physician's advice on acute myocardial infarction (symptoms and how to respond) (n = 1) [20]	History of myocardial infarction in a family (n = 1) [44]
Previous encounter with EMS system/experience in calling EMS (n = 2) [27,43]	Presence of ≥ 1 cardiovascular disease risk factor (n = 1) [20]
Utilization of EMS to give birth in a health facility in the previous 12 months (n = 1) [51]	Risk of myocardial infarction (n = 1) [44]
Self-reported probability of calling EMS in case of emergency (n = 1) [46]	Previous use of an ambulance for transportation to a hospital (n = 1) [24]
Self-reported willingness to call EMS for worsening symptoms of COVID-19 (n = 1) [49]	

EMS, emergency medical services; CPR, cardiopulmonary resuscitation.

recommendations regarding how to enhance people's knowledge of emergency phone numbers [12,13,17,18,20–23,26–28,30,31,33–36,38,40–44,46–55,57]. These included the suggestion to implement targeted or community-wide educational interventions (in particular, mandatory training) and public awareness campaigns about how to access EMS, as well as to institute and promote a unified national or international emergency number.

DISCUSSION

To the best of the author's knowledge, this review represents the first attempt to summarize and systematically map cumulative evidence from studies that examined public knowledge of emergency phone numbers and factors associated with this knowledge.

The percentage of people who knew emergency phone numbers varied significantly across the reviewed studies. In part, this can be explained by actual differences in public knowledge between different countries and regions, where high-income countries showed approximately two times greater community awareness of emergency numbers than countries with low or middle income. However, the reviewed research demonstrated substantial methodological heterogeneity. Different study designs, populations, sample sizes, geographic levels, and methods of data collection have undoubtedly affected the reported results. This suggests the need for developing uniform Utstein-style guidelines on the methodology of research and reporting of public awareness about how to respond to health emergencies in order to reduce possible bias associated with inconsistent research techniques and outcome measures.

Notwithstanding the acknowledged methodological differences, the reviewed studies detected alarmingly low community knowledge of emergency phone numbers and imply that this problem has a global scale. Based on the aggregate published data, approximately 3 out of 10 people in high-income countries and 6 out of 10 people in the low- and middle-income countries would be unable to call EMS because they do not know what number to dial. Even though several studies showed a relatively high percentage of people who knew what number to call in an emergency (sometimes exceeding 95% [10,52]), the non-absolute community awareness of how to phone EMS indicates the high probability of loss of life because of the bystanders' failure to activate emergency response in time-dependent health crises. Furthermore, reporting the correct emergency phone number as part of a survey does not confirm the respondent's ability to call EMS in a real-life emergency [21]. As demonstrated by Maes et al. [29], whereas 54.1% of study participants reported the correct emergency response number when completing a baseline questionnaire, only 28.2% dialed the correct number in a simulated scenario of sudden cardiac arrest.

Studies that attempted to examine factors associated with public awareness of emergency phone numbers have demonstrated contradictory results. For instance, although most research showed that subject age and gender are not associated with knowledge of EMS numbers, several studies observed an as-

sociation with different age and gender attributes of the participants. This suggests that determinants of community awareness of emergency phone numbers may have country- and region-specific patterns. Hence, detailed investigation of multiple factors is advisable for future studies, as the results may guide the development of targeted interventions to improve community preparedness for emergency response [7].

Four studies assessed the association of knowledge of emergency numbers with previous first aid or cardiopulmonary resuscitation training, and a positive relationship was identified in all instances. This can partly explain the considerably lower awareness of emergency phone numbers among populations from low- and middle-income countries where the availability of community training on life-saving principles and techniques is much lower than in the developed world [59]. Other major contributors to low public knowledge of emergency phone numbers in developing countries could be poor availability of EMS (including lack of a country-wide organized EMS system, decentralized EMS, insufficient EMS resources, and fee-based EMS), lack of community awareness of the need and purpose for EMS, and cultural issues (e.g., people's belief and use of traditional healing methods) [31,46,60,61]. It is worthy of note that studies that reported the lowest percentage of subjects who knew emergency numbers, namely 3.4% in Accra (Ghana) [31], 3.0% in Baghdad (Iraq) [22], and 0% in Libreville (Gabon) [23], were surveys conducted in capital cities of middle-income countries where organized EMS systems with universal emergency phone numbers were made available to the public before the studies' commencement. This indicates the importance of accompanying the establishment and development of new EMS systems with large-scale information campaigns to enhance community awareness of the offered services, their benefits, and the optimal ways to access EMS.

Whereas international guidelines acknowledge the essential role of early activation of EMS as part of bystander response to life-threatening emergencies [1,2], the focus of the professional community on the problem of insufficient public awareness of emergency phone numbers is generally low. The current review found that for most countries of the world, there is a lack of information on the extent of public knowledge of emergency numbers. The absence of this information could be a major limitation to system-level optimization of prehospital management for time-dependent medical events, as the integrity of the first (early access) link in the chain of survival is essential to the activation and viability of the whole sequence of care. High-quality national- and regional-level studies on the current topic seek to meet the needs of governments, public health, and EMS stakeholders. Beyond the paucity of research evaluating existing knowledge of emergency numbers and associated factors, few studies to date have explored the development and evaluation of methods to improve laypeople's ability to access EMS. A call to action from authoritative international stakeholders is necessary, with a focus on recognizing and emphasizing the research gap, incorporating this issue into international research agendas, and potentiating research aimed at exploring the best practices, methods, and tools for enhancing public awareness of EMS access. Considering the profound vulnerability of populations in the developing world, optimal research, education and awareness strategies tailored to low-resource settings is a priority area, along with the establishment and improvement of publicly accessible organized EMS systems in low- and middle-income countries. To avoid unreasonable overutilization of EMS, education and awareness campaigns should also inform the community of how to identify a health problem as an emergency case and when it is appropriate to activate EMS.

Limitations

The current study has limitations. In particular, the review did not consider publications other than peer-reviewed articles published in English in the specified 20-year timeframe. Some eligible studies could have been omitted due to the use of only one literature database, although efforts were made to ensure maximum coverage of eligible articles by searching grey literature and reference lists. Studies where knowledge of emergency phone numbers was examined but not reported among the main results in the abstract could also have been missed. Furthermore, the observational nature and the high methodological heterogeneity of the analyzed research warrant cautious interpretation of the results of this scoping review. Finally, the whole review procedure was completed by one author, and this may have introduced bias.

Conclusions

Based on the available studies, a large proportion of the public does not know how to phone EMS and therefore would be unable to activate an emergency response in time-dependent health crises. The prevalence of this issue in low- and middle-income countries is particularly concerning and warrants further investigation. The non-absolute public knowledge of how to activate EMS constitutes a risk factor for potentially preventable deaths and disabilities. Greater organizational and high-quality research efforts are needed to better understand the current extent and nature of the problem and to develop effective solutions to enhance the general population's ability to access EMS.

ARTICLE INFORMATION

Conflicts of interest
The author has no conflicts of interest to declare.

Funding
The author received no financial support for this study.

Data availability
Data analyzed in this study are openly available from the Mendeley Data repository at https://doi.org/10.17632/s4pyb9k2w8.1.

Supplementary materials
Suppl. 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
Suppl. 2. Search strategy for PubMed.
Supplementary materials are available from https://doi.org/10.15441/ceem.24.289.

REFERENCES

1. Berg KM, Cheng A, Panchal AR, et al. Part 7: Systems of care: 2020 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2020;142(16_suppl_2): S580–604.
2. Greif R, Lockey A, Breckwoldt J, et al. European Resuscitation Council guidelines 2021: education for resuscitation. Resuscitation 2021;161:388–407.
3. Cabrita B, Bouyer-Dalloz F, L'Huillier I, et al. Beneficial effects of direct call to emergency medical services in acute myocardial infarction. Eur J Emerg Med 2004;11:12–8.
4. Nehme Z, Andrew E, Cameron P, et al. Direction of first bystander call for help is associated with outcome from out-of-hospital cardiac arrest. Resuscitation 2014;85:42–8.
5. Duvekot MH, Kerkhoff H, Venema E, et al. Medical attention seeking by suspected stroke patients: emergency medical services or general practitioner? Clin Neurol Neurosurg 2022; 218:107297.
6. Eisenberg M, Hallstrom A, Becker L. Community awareness of emergency phone numbers. Am J Public Health 1981;71: 1058–60.
7. Clark MJ, Enraght-Moony E, Balanda KP, Lynch M, Tighe T, FitzGerald G. Knowledge of the national emergency telephone number and prevalence and characteristics of those trained in CPR in Queensland: baseline information for targeted training interventions. Resuscitation 2002;53:63–9.
8. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169:467–73.
9. Birkun A. Dataset of original studies that reported public knowledge of telephone numbers for medical emergencies [Internet]. Version 1. Mendeley Data; 2024 [cited 2024 Jul 19]. Available from: https://doi.org/10.17632/s4pyb9k2w8.1
10. Connolly M, Toner P, Connolly D, McCluskey DR. The 'ABC for life' programme: teaching basic life support in schools. Resuscitation 2007;72:270–9.
11. Baez AA, Giraldez E, Lane PL, Pozner C, Rodriguez J, Rogers S. Knowledge and attitudes of the out-of-hospital emergency care consumers in Santo Domingo, Dominican Republic. Prehosp Disaster Med 2008;23:373–6.
12. Pergola AM, Araujo IE. The layperson in emergency situation. Rev Esc Enferm USP 2008;42:769–76.
13. Pontes-Neto OM, Silva GS, Feitosa MR, et al. Stroke awareness in Brazil: alarming results in a community-based study. Stroke 2008;39:292–6.
14. Bollig G, Wahl HA, Svendsen MV. Primary school children are able to perform basic life-saving first aid measures. Resuscitation 2009;80:689–92.
15. Jennings S, Hara TO, Cavanagh B, Bennett K. A national survey of prevalence of cardiopulmonary resuscitation training and knowledge of the emergency number in Ireland. Resuscitation 2009;80:1039–42.
16. Kirschner K, Braspenning J, Maassen I, Bonte A, Burgers J, Grol R. Improving access to primary care: the impact of a quality-improvement strategy. Qual Saf Health Care 2010;19: 248–51.
17. Rajapakse R, Noc M, Kersnik J. Public knowledge of cardiopulmonary resuscitation in Republic of Slovenia. Wien Klin Wochenschr 2010;122:667–72.
18. Subramaniam MR, Mahajan PV, Knazik SR, Giblin PT, Thomas R, Kannikeswaran N. Awareness and utilization of emergency medical services by limited English proficient caregivers of pediatric patients. Prehosp Emerg Care 2010;14:531–6.
19. Kanstad BK, Nilsen SA, Fredriksen K. CPR knowledge and attitude to performing bystander CPR among secondary school students in Norway. Resuscitation 2011;82:1053–9.
20. Kopec G, Sobien B, Podolec M, et al. Knowledge of a patient-dependant phase of acute myocardial infarction in Polish adults: the role of physician's advice. Eur J Public Health 2011;21:603–8.
21. Piepho T, Resch N, Heid F, Werner C, Noppens RR. Lay basic

life support: the current situation in a medium-sized German town. Emerg Med J 2011;28:786–9.
22. Shanovich PK, Donaldson RI, Hung YW, Hasoon T, Evans GE. Iraqi community members' knowledge, attitude and practice of emergency medical care: assessing civilian emergency medicine in an area of conflict. Med Confl Surviv 2011;27:151–64.
23. Bosson N, Redlener MA, Foltin GL, Raven MC, Foran MP, Wall SP. Barriers to utilization of pre-hospital emergency medical services among residents in Libreville, Gabon: a qualitative study. Afr J Emerg Med 2013;3:172–7.
24. Demirkan B, Ege MR, Dogan P, Ipek EG, Guray U, Guray Y. Factors influencing the use of ambulance among patients with acute coronary syndrome: results of two centers in Turkey. Anadolu Kardiyol Derg 2013;13:516–22.
25. Goktas S, Yildirim G, Kose S, Yildirim S, Ozhan F, Senturan L. First aid knowledge of university students in poisoning cases. Turk J Emerg Med 2016;14:153–9.
26. Lim W, Chuang DF, Chue KM, et al. Stroke literacy in Singapore: data from a survey of public housing estate residents. Ann Acad Med Singap 2014;43:454–63.
27. EMSAT: The Emergency Medical Service Assessment Team group; Hamam AF, Bagis MH, AlJohani K, Tashkandi AH. Public awareness of the EMS system in Western Saudi Arabia: identifying the weakest link. Int J Emerg Med 2015;8:70.
28. Kopec G, Waligora M, Brozda M, et al. Effectiveness of single medical advice on emergency phone number knowledge in urban adult population: "Health, Alcohol and Psychosocial Factors in Eastern Europe" substudy. Przegl Epidemiol 2015;69:543-8, 649-52.
29. Maes F, Marchandise S, Boileau L, Le Polain de Waroux JB, Scavee C. Evaluation of a new semiautomated external defibrillator technology: a live cases video recording study. Emerg Med J 2015;32:481–5.
30. Mobarak AS, Afifi RM, Qulali A. First aid knowledge and attitude of secondary school students in Saudi Arabia. Health 2015;7:1366–78.
31. Mould-Millman NK, Rominski SD, Bogus J, et al. Barriers to accessing emergency medical services in Accra, Ghana: development of a survey instrument and initial application in Ghana. Glob Health Sci Pract 2015;3:577–90.
32. Alhejaili AS, Alsubhi SA. Knowledge and attitude of first aid skills among health science students at Taibah University. J Gen Pract 2016;4:257.
33. Callachan EL, Alsheikh-Ali AA, Nair SC, Bruijns S, Wallis LA. Utilizations and perceptions of emergency medical services by patients with ST-segments elevation acute myocardial infarction in Abu Dhabi: a multicenter study. Heart Views 2016;17:49–54.
34. Banfai B, Pek E, Pandur A, Csonka H, Betlehem J. 'The year of first aid': effectiveness of a 3-day first aid programme for 7-14-year-old primary school children. Emerg Med J 2017;34:526–32.
35. Brinkrolf P, Bohn A, Lukas RP, et al. Senior citizens as rescuers: Is reduced knowledge the reason for omitted lay-resuscitation-attempts?: results from a representative survey with 2004 interviews. PLoS One 2017;12:e0178938.
36. Einav S, Wacht O, Kaufman N, Alkalay E. Cardiopulmonary arrest in primary care clinics: more holes than cheese: a survey of the knowledge and attitudes of primary care physicians regarding resuscitation. Isr J Health Policy Res 2017;6:22.
37. Banfai B, Pandur A, Schiszler B, et al. Little lifesavers: can we start first aid education in kindergarten?: a longitudinal cohort study. Health Educ J 2018;77:1007–17.
38. Modi PD, Solanki R, Nagdev TS, et al. Public awareness of the emergency medical services in Maharashtra, India: a questionnaire-based survey. Cureus 2018;10:e3309.
39. Mathiesen WT, Birkenes TS, Lund H, Ushakova A, Soreide E, Bjorshol CA. Public knowledge and expectations about dispatcher assistance in out-of-hospital cardiac arrest. J Adv Nurs 2019;75:783–92.
40. Midani O, Tillawi T, Saqer A, Hammami MB, Taifour H, Mohammad H. Knowledge and attitude toward first aid: a cross-sectional study in the United Arab Emirates. Avicenna J Med 2019;9:1–7.
41. Sri-On J, Vanichkulbodee A, Sinsuwan N, Rojsaengroeng R, Kamsom A, Liu SW. Disaster preparedness among Thai elderly emergency department patients: a survey of patients' perspective. BMC Emerg Med 2019;19:58.
42. Sultan M, Abebe Y, Tsadik AW, Ababa A, Yesus AG, Mould-Millman NK. Trends and barriers of emergency medical service use in Addis Ababa; Ethiopia. BMC Emerg Med 2019;19:28.
43. Alabdali AA, Alfraidi AA, Almuhanna AA, et al. Public awareness of emergency medical services phone number. Saudi J Emerg Med 2021;2:147–52.
44. Korman M, Felkle D, Korman T. Are rural patients aware of myocardial infarction's risk factors, symptoms, and management?: a survey study in the COVID-19 pandemic era. Folia Med Cracov 2021;61:103–15.
45. Sturny L, Regard S, Larribau R, et al. Differences in basic life support knowledge between junior medical students and lay

people: web-based questionnaire study. J Med Internet Res 2021;23:e25125.
46. Takele GM, Ballo TH, Gebrekidan KB, Gebregiorgis BG. Utilization, barriers and determinants of emergency medical services in Mekelle City, Tigray, Ethiopia: a community-based cross-sectional study. Open Access Emerg Med 2021;13:325–34.
47. Aljabri D, Albinali H. Public awareness and use of 997 emergency medical service phone number during the COVID-19 pandemic. Front Public Health 2022;10:937202.
48. Durr D, Carron PN, Ageron FX, et al. Basic life support knowledge among Swiss conscripts: a national multicentre survey. Swiss Med Wkly 2022;152:w30147.
49. Olani A, Beza Demisse L, Alemayehu M, Sultan M. Factors associated with willingness to call emergency medical services for worsening symptoms of COVID-19 in Addis Ababa, Ethiopia. Open Access Emerg Med 2022;14:563–72.
50. Rodrigues MS, Santana LF, Fernandes Castro AP, et al. Awareness towards stroke among high school students in Brazil: a cross-sectional study. Sao Paulo Med J 2022;140:525–30.
51. Tolossa W, Bala ET, Mekuria M, Ifa M, Deriba BS, Dufera A. Magnitude and factors associated with ambulance service utilization among women who gave birth at public health institutions in central Ethiopia. Open Access Emerg Med 2022;14:457–71.
52. von dem Knesebeck O, Koens S, Schafer I, Straub A, Klein J. Public knowledge about emergency care: results of a population survey from Germany. Front Public Health 2022;9:787921.
53. Wongtanasarasin W, Sutham K, Laosuksri W, Wittayachamnankul B. Health characteristics, knowledge, and attitude towards basic life support among marathon runners in Thailand: a population-based survey. J Sports Med Phys Fitness 2022;62:1742–7.
54. Gomes OV, Guimaraes MP, Barbosa BM, et al. Awareness of stroke among patients with chronic kidney disease on hemodialysis: a cross-sectional study. Sao Paulo Med J 2023;142:e2022644.
55. Moreno AP, Camargo L, Gaitan G, et al. Effectiveness of a digital application to improve stroke knowledge for kids. Neurologia (Engl Ed) 2023;38:278–83.
56. Pedrazas-Lopez D, de Pablo-Marquez B, Cunillera-Puertolas O, Almeda-Ortega J; Grupo de Investigacion RCParvulari. RCParvulari training: a basic life support training methodology applied to 5-year-old students: effectiveness in a cluster-randomized clinical trail. An Pediatr (Engl Ed) 2023;98:99–108.
57. Mohamadiyeh A, Said MR, Al-Dabbagh B, et al. Knowledge of and attitude towards first aid measures among drivers in Sharjah. Emerg Care Med 2024;1:3–10.
58. The World Bank Group. World Bank country and lending groups [Internet]. The World Bank Group; 2024 [cited 2024 Jul 19]. Available from: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups
59. Birkun A, Gautam A, Trunkwala F. Global prevalence of cardiopulmonary resuscitation training among the general public: a scoping review. Clin Exp Emerg Med 2021;8:255–67.
60. Mould-Millman NK, Dixon JM, Sefa N, et al. The state of emergency medical services (EMS) systems in Africa. Prehosp Disaster Med 2017;32:273–83.
61. Plummer V, Boyle M. EMS systems in lower-middle income countries: a literature review. Prehosp Disaster Med 2017;32:64–70.

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):37-43
https://doi.org/10.15441/ceem.25.055

Original Article

pISSN: 3022-1811
eISSN: 2383-4625

Epidemiologic trends in pediatric urinary tract infections in United States emergency departments from 2016 to 2023

Michael Gottlieb[1], Tamara Amponsah[1], Nhat Nguyen[2], Ohm Shukla[2], Kyle Bernard[1], Eric Moyer[1]

[1]Department of Emergency Medicine, Rush University Medical Center, Chicago, IL, USA
[2]Rush Medical College, Chicago, IL, USA

Objective Urinary tract infections are a common consideration when assessing pediatric patients with fever. With rising resistance rates and increased focus on antibiotic stewardship, current management needs to be better understood. This study reports the incidence and antibiotic distribution in a nationwide cohort of pediatric emergency department (ED) patients with cystitis and pyelonephritis during an 8-year period.

Methods We performed a cohort study from January 1, 2016, to December 31, 2023, using Epic Cosmos data. Pediatric ED patients (<18 years) with an ICD-10 code corresponding to cystitis or pyelonephritis were included. The outcomes were the total number of ED presentations, outpatient antibiotic prescriptions, and antibiotics administered in the ED for patients admitted with cystitis or pyelonephritis. Binary logistic regression models were used to measure the relationship between the year and the dependent variables.

Results Among 46,774,814 total pediatric ED visits, 720,863 (1.5%) were for cystitis, and 82,717 (0.2%) were for pyelonephritis. Among those admitted, the most common antibiotics were third-generation cephalosporins (cystitis, 55.6%; pyelonephritis, 62.3%), first-generation cephalosporins (cystitis, 13.8%; pyelonephritis, 13.7%), and ampicillin (cystitis, 10.8%; pyelonephritis, 6.6%). First-generation cephalosporin use rose over time, while ampicillin and ciprofloxacin use declined. Among discharged patients, the most common antibiotics were first-generation cephalosporins (cystitis, 43.1%; pyelonephritis, 33.7%), third-generation cephalosporins (cystitis, 20.8%; pyelonephritis, 25.8%), and trimethoprim-sulfamethoxazole (cystitis, 13.5%; pyelonephritis, 11.7%). First-generation cephalosporin use rose over time, while trimethoprim-sulfamethoxazole and ciprofloxacin use declined. Among those with cystitis specifically, the use of third-generation cephalosporins declined over time.

Conclusion Cystitis and pyelonephritis remain common ED presentations, representing nearly 2% of all pediatric ED visits, and antibiotic selection has shifted notably over time. Understanding the current epidemiology can inform public health planning and antibiotic stewardship in the ED.

Keywords Epidemiology; Urinary tract infection; Cystitis; Pyelonephritis; Pediatric

Received: March 4, 2025
Revised: April 22, 2025
Accepted: April 23, 2025

Correspondence to:
Michael Gottlieb
(MichaelGottliebMD@gmail.com)

How to cite this article:
Gottlieb M, Amponsah T, Nguyen N, Shukla O, Bernard K, Moyer E. Epidemiologic trends in pediatric urinary tract infections in United States emergency departments from 2016 to 2023.
Clin Exp Emerg Med 2026;13(1):37-43.
https://doi.org/10.15441/ceem.25.055

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

© 2026 The Korean Society of Emergency Medicine

Capsule Summary

What is already known

Pediatric urinary tract infections are a common reason for presentation to the emergency department.

What is new in the current study

From 2016–2023, urinary tract infections accounted for nearly 2% of all pediatric emergency department visits, and the most common antibiotics varied by cystitis versus pyelonephritis and admission versus discharge.

INTRODUCTION

Urinary tract infections (UTIs) are a common consideration in the differential diagnosis of fever because they are among the most common bacterial infections in the pediatric population [1]. Among infants presenting with fever, UTI can be present in up to 7% of patients [2]. UTIs remain common throughout childhood, with a pooled prevalence of UTI in the pediatric population of 7.8% [2].

UTIs vary in severity, ranging from cystitis (an infection localized to the bladder) to pyelonephritis (extension to the kidneys, often requiring different and longer courses of antibiotics). However, most studies have combined these conditions, limiting the ability to delineate the prevalence of cystitis and pyelonephritis [2,3]. Moreover, UTIs are a leading cause of antibiotic prescriptions for pediatric patients and a major risk for antibiotic resistance [3]. Therefore, it is critical to understand the current epidemiology of both cystitis and pyelonephritis and the current antibiotic usage trends to inform healthcare planning and antibiotic stewardship efforts.

Updated research is thus needed to understand the treatments being prescribed for pediatric cystitis and pyelonephritis in emergency departments (EDs). To meet that need, we examined the current incidence and antibiotic prescribing practices for cystitis and pyelonephritis in EDs.

METHODS

Ethics statement

The Institutional Review Board of Rush University Medical Center deemed this study as nonhuman subject research and waived the requirement for informed consent. This study adheres to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting guidelines [4].

Study design and setting

We conducted a cohort study of pediatric visits to United States EDs for cystitis and pyelonephritis and the associated treatments and dispositions across an 8-year period using the Cosmos research platform (Epic Systems Corp). Cosmos is a large database that aggregates electronic health record data voluntarily submitted by health systems for the purposes of research. Cosmos data are representative of the United States population and consistent with national census data. At the time of submission, the Cosmos database included more than 289 million patients across 39,000 hospitals and clinics. Patients with records at multiple institutions are deduplicated and anonymized centrally by Epic.

We performed queries of Cosmos using the International Classification of Diseases, 10th Revision (ICD-10) codes corresponding to cystitis (Suppl. 1) and pyelonephritis (Suppl. 2) from January 1, 2016, to December 31, 2023. We selected this time period based on the launch of ICD-10 coding requirements in October 2015. The inclusion criteria were pediatric patients (< 18 years) with an ED presentation within the United States and an associated ICD-10 code corresponding to cystitis or pyelonephritis.

Statistical analysis

We analyzed the incidence of cystitis and pyelonephritis presentations, antibiotic treatment, and disposition. Antibiotics were categorized based on their classes and separately analyzed for admitted and discharged patients. Discharge antibiotics were based on prescriptions. Analyses were further subgrouped by age (< 2 years vs. ≥ 2 years) to account for differences in antibiotic prescribing. We used binary logistic regression models to measure the relationship between the year and the dependent variables (i.e., ED encounters due to cystitis, ED encounters due to pyelonephritis, antibiotics). Data are reported as odds ratios (ORs) with 95% confidence intervals (CIs). Analyses were performed with IBM SPSS ver. 26.0 (IBM Corp).

RESULTS

Of 46,774,814 ED encounters with pediatric patients during the 8-year period, 720,863 (1.5%) were due to cystitis, and 82,717 (0.2%) were due to pyelonephritis, with both diseases experiencing a slight rise over time (cystitis: OR, 1.006 [95% CI, 1.005–1.007]; pyelonephritis: OR, 1.019 [95% CI, 1.016–1.022]) (Suppls. 3–8). The highest rate was seen in 2020, when cystitis was present in 1.7% of visits and pyelonephritis in 0.2%. The demographics of the study populations with cystitis and pyelonephritis are reported in Tables 1 and 2.

For patients with cystitis who were admitted, the most common antibiotic class was third-generation cephalosporins (55.6%), followed by first-generation cephalosporins (13.8%) and ampicillin (10.8%). When evaluated over time, there was a notable rise in the use of first-generation cephalosporins (8.3% to 18.5%; OR, 1.157 [95% CI, 1.140–1.175]), fourth-generation cephalosporins (2.7% to 7.8%; OR, 1.146 [95% CI, 1.121–1.171]), meropenem (2.1% to 4.7%; OR, 1.107 [95% CI, 1.077–1.138]), and ertapenem (0.5% to 1.3%; OR, 1.149 [95% CI, 1.085–1.217]) (Fig. 1A, Suppl. 9). There was a decline in the use of third-generation cephalosporins (59.2% to 52.6%; OR, 0.962 [95% CI, 0.953–0.972]), ampicillin (13.0% to 9.4%; OR, 0.953 [95% CI, 0.938–0.968]), ciprofloxacin (4.0% to 2.6%; OR, 0.954 [95% CI, 0.929–0.981]), and second-generation cephalosporins (1.1% to 0.5%; OR, 0.864 [95% CI, 0.811–0.920]). When analyzed by age (<2 years vs. ≥2 years), antibiotic selection was similar overall; however, ampicillin (18.0% vs. 3.8%) and gentamicin (9.6% vs. 4.0%) were more commonly prescribed to those <2 years, while fourth-generation cephalosporins (4.4% vs. 8.2%) and piperacillin-tazobactam (1.3% vs. 5.6%) were more commonly prescribed to those ≥2 years (Suppls. 10, 11).

Table 1. Demographics of the pediatric patients (<18 years) with cystitis

Demographic	No. of patients (%)		
	Total (n = 720,863)	<2 yr (n = 101,120)	≥2 yr (n = 619,743)
Sex			
Female	648,497 (90.0)	73,350 (72.5)	575,147 (92.8)
Male	72,280 (10.0)	27,764 (27.5)	44,516 (7.2)
Other/not reported	86 (0.0)	6 (0.0)	80 (0.0)
Race[a]			
American Indian/Alaskan Native	10,367 (1.4)	1,357 (1.3)	9,010 (1.5)
Asian	18,985 (2.6)	4,826 (4.8)	14,159 (2.3)
Black/African American	181,269 (25.1)	19,142 (18.9)	162,127 (26.2)
Native Hawaiian/other Pacific Islander	5,769 (0.8)	1,198 (1.2)	4,571 (0.7)
White	450,131 (62.4)	62,217 (61.5)	387,914 (62.6)
Other	138,336 (19.2)	22,601 (22.4)	115,735 (18.7)
Not reported	18,143 (2.5)	3,879 (3.8)	14,264 (2.3)
Ethnicity			
Hispanic or Latino	173,812 (24.1)	32,546 (32.2)	141,266 (22.8)
Other	507,086 (70.3)	62,314 (61.6)	444,772 (71.8)
Not reported	39,965 (5.5)	6,260 (6.2)	33,705 (5.4)
Insurance[a]			
Private/other	412,324 (57.2)	58,736 (58.1)	353,588 (57.1)
Medicaid	298,214 (41.4)	42,077 (41.6)	256,137 (41.3)
Medicare	603 (0.1)	113 (0.1)	490 (0.1)
Self-pay	23,973 (3.3)	1,698 (1.7)	22,275 (3.6)
Not reported	12,722 (1.8)	1,343 (1.3)	11,379 (1.8)
US Census Bureau–designated region			
South	375,914 (52.1)	53,654 (53.1)	322,260 (52.0)
Midwest	188,226 (26.1)	20,366 (20.1)	167,860 (27.1)
Northeast	85,035 (11.8)	13,009 (12.9)	72,026 (11.6)
West	70,835 (9.8)	13,933 (13.8)	56,902 (9.2)
Not reported	853 (0.1)	158 (0.2)	695 (0.1)

[a]The subtotal exceeds the total because some individuals selected more than one.

Table 2. Demographics of pediatric patients (<18 years) with pyelonephritis

Demographic	No. of patients (%)		
	Total (n = 82,717)	<2 yr (n = 12,130)	≥2 yr (n = 70,587)
Sex			
Female	73,660 (89.1)	8,537 (70.4)	65,123 (92.3)
Male	9,041 (10.9)	3,593 (29.6)	5,448 (7.7)
Other/not reported	16 (0.0)	0 (0)	16 (0.0)
Race[a]			
American Indian or Alaskan Native	1,293 (1.6)	176 (1.5)	1,117 (1.6)
Asian	2,458 (3.0)	781 (6.4)	1,677 (2.4)
Black or African American	14,074 (17.0)	1,888 (15.6)	12,186 (17.3)
Native Hawaiian or other Pacific Islander	764 (0.9)	199 (1.6)	565 (0.8)
White	59,107 (71.5)	7,548 (62.2)	51,559 (73.0)
Other	14,518 (17.6)	2,853 (23.5)	11,665 (16.5)
Not reported	1,855 (2.2)	427 (3.5)	1,428 (2.0)
Ethnicity			
Hispanic or Latino	17,997 (21.8)	4,052 (33.4)	13,945 (19.8)
Other	60,163 (72.7)	7,316 (60.3)	52,847 (74.9)
Not reported	4,557 (5.5)	762 (6.3)	3,795 (5.4)
Insurance[a]			
Private or other	51,770 (62.6)	7,276 (60.0)	44,494 (63.0)
Medicaid	30,640 (37.0)	4,784 (39.4)	25,856 (36.6)
Medicare	85 (0.1)	16 (0.0)	69 (0.1)
Self-pay	2,336 (2.8)	152 (1.3)	2,184 (3.1)
Not reported	1,482 (1.8)	243 (2.0)	1,239 (1.8)
US Census Bureau–designated region			
South	37,583 (45.4)	5,962 (49.2)	31,621 (44.8)
Midwest	21,204 (25.6)	2,053 (16.9)	19,151 (27.1)
Northeast	11,563 (14.0)	1,782 (14.7)	9,781 (13.9)
West	12,282 (14.8)	2,315 (19.1)	9,967 (14.1)
Not reported	85 (0.1)	18 (0.1)	67 (0.1)

[a]The subtotal exceeds the total because some individuals selected more than one.

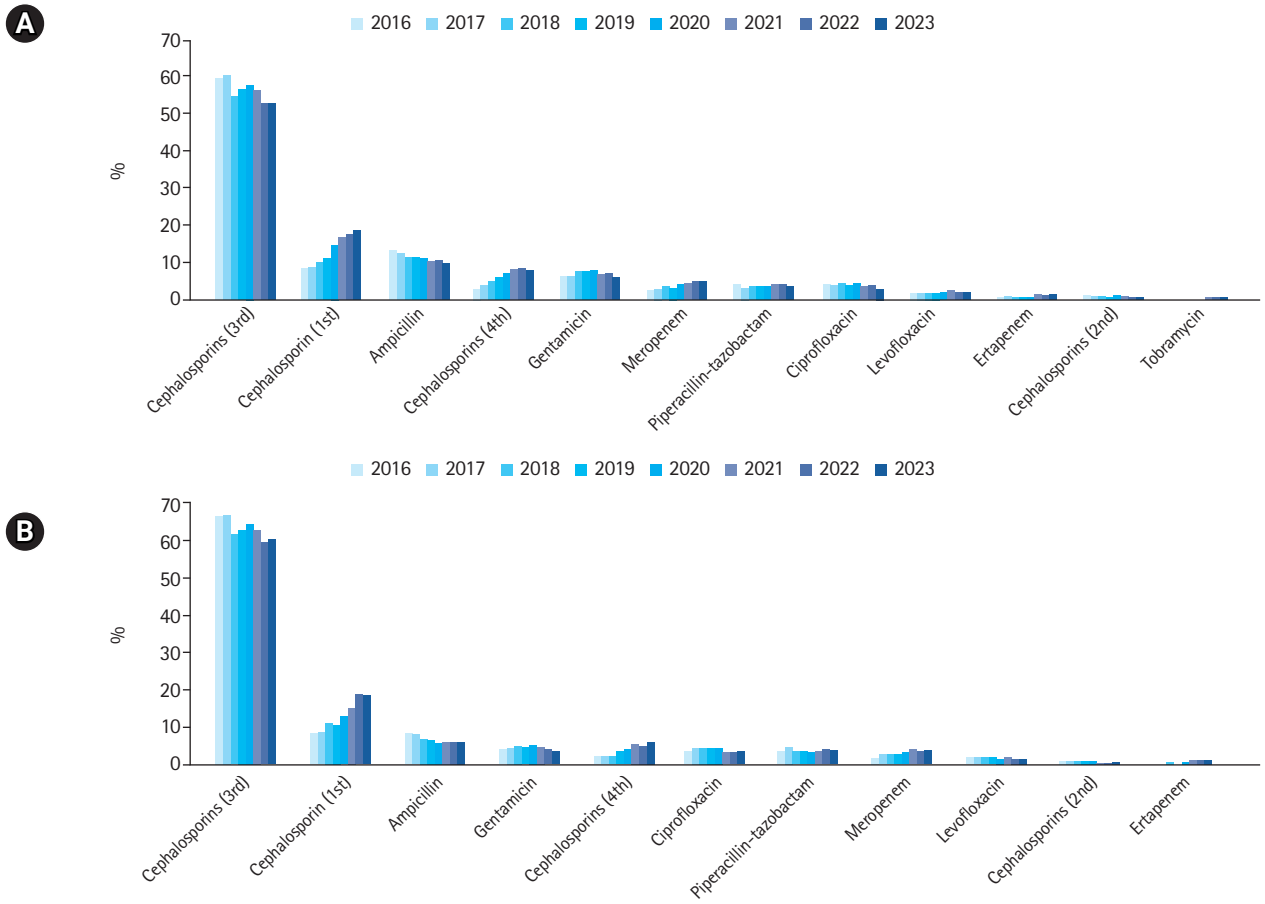


Fig. 1. Antibiotic medication classes administered to pediatric patients in the emergency department (2016–2023). (A) Admitted with cystitis. (B) Admitted with pyelonephritis.

For patients with pyelonephritis who were admitted, the most common antibiotic class was third-generation cephalosporins (62.3%), followed by first-generation cephalosporins (13.7%) and ampicillin (6.6%). When evaluated over time, there was a notable rise in the use of first-generation cephalosporins (8.4% to 13.7%; OR, 1.155 [95% CI, 1.132–1.178]), fourth-generation cephalosporins (2.1% to 6.1%; OR, 1.186 [95% CI, 1.145–1.228]), and meropenem (2.1% to 3.8%; OR, 1.085 [95% CI, 1.045–1.127]) (Fig. 1B, Suppl. 12). There was a decline in the use of third-generation cephalosporins (66.3% to 59.9%; OR, 0.963 [95% CI, 0.950–0.976]), ampicillin (8.4% to 6.0%; OR, 0.947 [95% CI, 0.923–0.971]), and second-generation cephalosporins (1.1% to 0.8%; OR, 0.881 [95% CI, 0.820–0.946]). When analyzed by age (<2 years vs. ≥2 years), antibiotic selection was similar overall; however, ampicillin (14.0% vs. 3.8%) and gentamicin (6.2% vs. 3.7%) were more commonly prescribed to those <2 years, while piperacillin-tazobactam (1.0% vs. 4.6%) was more commonly prescribed to those ≥2 years (Suppls. 13, 14).

Among discharged patients with cystitis, the most common antibiotic class was first-generation cephalosporins (43.1%), followed by third-generation cephalosporins (20.8%), trimethoprim-sulfamethoxazole (13.5%), and nitrofurantoin (7.8%). When evaluated over time, there was a notable rise in the use of first-generation cephalosporins (29.4% to 50.7%; OR, 1.144 [95% CI, 1.141–1.146]) (Fig. 2A, Suppl. 15). The use of most other antibiotics declined, including trimethoprim-sulfamethoxazole (22.4% to 8.4%; OR, 0.841 [95% CI, 0.838–0.843]), nitrofurantoin (8.9% to 6.7%; OR, 0.947 [95% CI, 0.944–0.951]), amoxicillin-clavulanate (4.3% to 2.8%; OR, 0.953 [95% CI, 0.948–0.958]), second-generation cephalosporins (3.0% to 1.7%; OR, 0.869 [95% CI, 0.863–0.876]), and ciprofloxacin (2.9% to 1.4%; OR, 0.841 [95% CI, 0.833–0.848]). When analyzed by age (<2 years vs. ≥2 years), antibiotic selection was similar overall; however, third-generation cephalosporins (33.1% vs. 19.1%) were more commonly prescribed to those <2 years, while trimethoprim-sulfamethoxazole (6.5% vs. 14.5%), and nitrofurantoin (0.8% vs. 8.9%) were more commonly prescribed to those ≥2 years (Suppls. 16, 17).

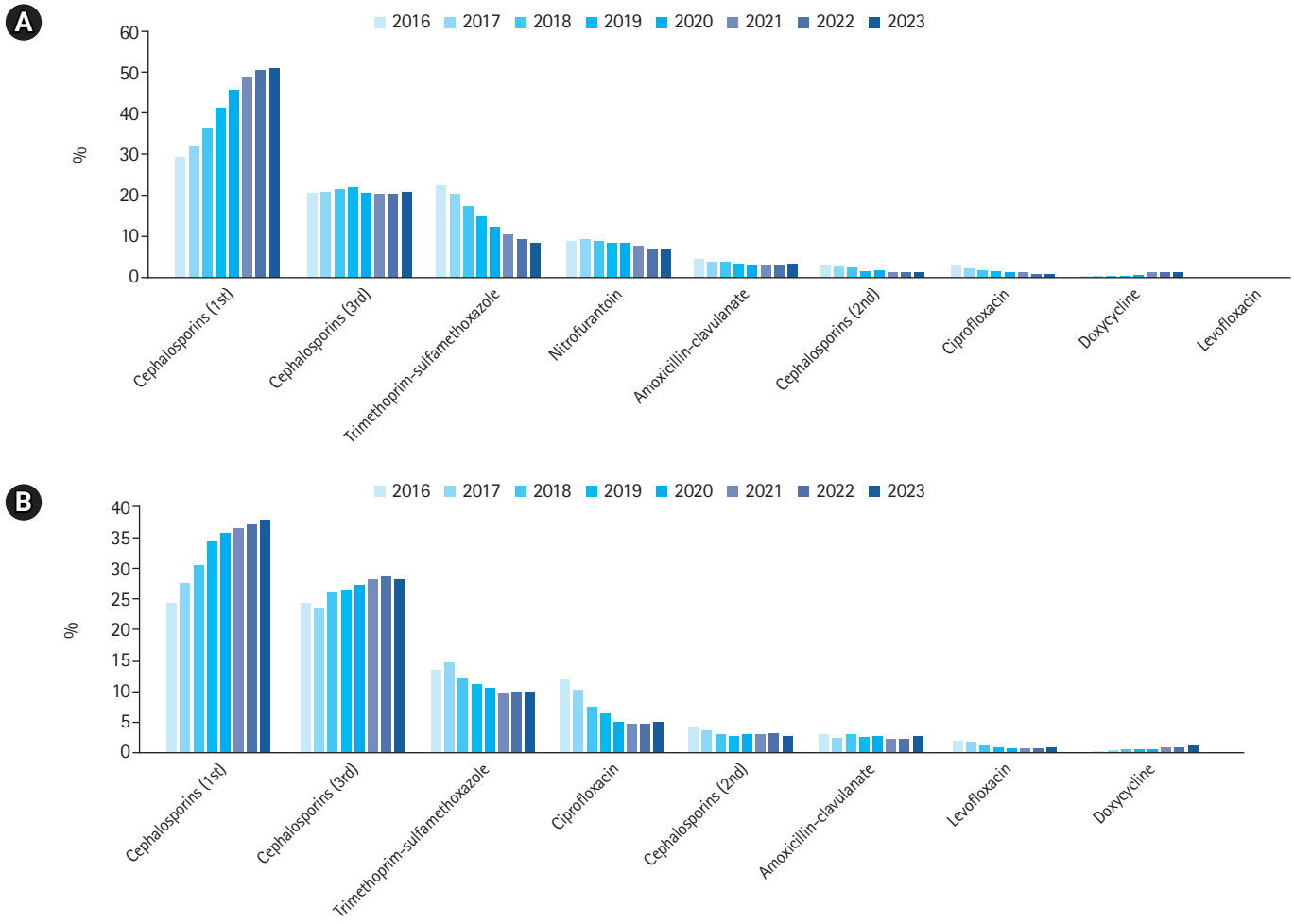


Fig. 2. Antibiotic prescriptions to pediatric patients in the emergency department (2016–2023). (A) Discharged with cystitis. (B) Discharged with pyelonephritis.

Because most children younger than 2 years with pyelonephritis are admitted, we focused the analyses of discharge prescriptions for pyelonephritis on those aged 2 to 17 years. For this population, the most common antibiotic was first-generation cephalosporins (33.7%), followed by third-generation cephalosporins (25.8%), trimethoprim-sulfamethoxazole (11.7%), and ciprofloxacin (6.9%) (Fig. 2B, Suppl. 18). When evaluated over time, there was a notable rise in the use of first-generation cephalosporins (24.2% to 36.5%; OR, 1.071 [95% CI, 1.062–1.079]) and third-generation cephalosporins (21.8% to 27.8%; OR, 1.055 [95% CI, 1.046–1.064]) and a decline in the use of trimethoprim-sulfamethoxazole (14.7% to 10.5%; OR, 0.934 [95% CI, 0.923–0.944]), ciprofloxacin (13.5% to 5.3%; OR, 0.852 [95% CI, 0.840–0.865]), and second-generation cephalosporins (3.9% to 2.9%; OR, 0.976 [95% CI, 0.956–0.997]).

DISCUSSION

In a large, nationally representative sample of more than 46 million pediatric visits to the ED, several key findings were identified. Overall, cystitis and pyelonephritis represented nearly 2% of all visits, with a slight rise over time. Interestingly, a spike in the incidence occurred in 2020, which coincided with a decrease in total pediatric ED visits. This might reflect decreased overall ED use during the COVID-19 pandemic, which thus resulted in a higher proportion of visits for UTIs without an absolute increase in the number of diagnoses.

Among those admitted for cystitis or pyelonephritis, the majority received a third-generation cephalosporin. This is consistent with common practice and guidelines from the American Academy of Pediatrics [5]. Interestingly, the use of third-generation cephalosporins has decreased over time, and the use of first-generation cephalosporins has increased. This might reflect success-

ful efforts to improve antibiotic stewardship because many pediatric UTIs are caused by *Escherichia coli*, which is often susceptible to first-generation cephalosporins and might not require broader-spectrum agents [6]. First-generation cephalosporins are also associated with a reduced risk of adverse events, including the development of *Clostridioides difficile* [7]. Similar declines were seen in trimethoprim-sulfamethoxazole prescriptions for cystitis, which might reflect rising resistance against this agent in pediatric UTIs [8].

Ciprofloxacin use has also declined. This finding is consistent with other national data showing reductions in fluoroquinolone use [9]. This shift could reflect recent data demonstrating rising resistance against fluoroquinolones by UTIs [10]. It might also be influenced by increased recognition of the adverse events associated with fluoroquinolones, including corrected QT interval prolongation and tendinopathy [11]. Moreover, the American Academy of Pediatrics has advised clinicians to avoid fluoroquinolones except for the treatment of infections for which no safe and effective alternative exists [12].

Although the use of narrow-spectrum antibiotics (e.g., first-generation cephalosporins) has risen in multiple patient populations, we also found a rise in the use of broader-spectrum antibiotics, such as meropenem, that are often reserved for multidrug resistant organisms [13]. This might reflect attempts to improve antibiotic stewardship with narrowing antibiotics when appropriate, but it could also reflect a rise in multidrug resistant organisms in general. Further research is needed to determine changes in resistance patterns and determine how antibiotic selection changes based on culture sensitivities.

Limitations

This study has several important limitations. Although the Cosmos database is representative of the United States population with regards to age, race, and ethnicity, it is limited to organizations that use Epic for their electronic health records and opt to contribute data to Cosmos. Thus, there is the potential for selection bias toward patients that seek care at such institutions, and the data might not fully reflect the entire cohort of pediatric patients with cystitis or pyelonephritis in the United States. Additionally, patients who were transferred to pediatric EDs that did not use Epic or were not part of the Cosmos database do not have available data on the antibiotics administered or prescribed. Moreover, ICD-10 codes were used to identify cases of cystitis and pyelonephritis. Relevant cases that were incorrectly coded might thus have been missed. Cases that were miscoded might also have led to incorrect crossover between cystitis and pyelonephritis. Finally, we were unable to separate out concomitant disease processes or other factors that might have influenced decisions about disposition or antibiotic selection.

Conclusions

Cystitis and pyelonephritis remain common ED presentations, representing nearly 2% of all pediatric ED visits with a slight rise over time. First- and third-generation cephalosporins remain the most commonly prescribed antibiotics, though there have been notable shifts in the prescribing of antibiotics over time. Understanding the current epidemiology can inform public health planning and antibiotic stewardship in the ED.

ARTICLE INFORMATION

Author contributions

Conceptualization: all authors; Investigation: all authors; Project administration: all authors; Supervision: MG; Visualization: all authors; Writing–original draft: all authors; Writing–review & editing: all authors. All authors read and approved the final manuscript.

Conflicts of interest

Michael Gottlieb is an editorial board member of this journal, but was not involved in the peer reviewer selection, evaluation, or decision process of this article. The authors have no other conflicts of interest to declare.

Funding

The authors received no financial support for this study.

Data availability

Data analyzed in this study are available within the article and its supplementary materials.

Supplementary materials

Suppl. 1. Cystitis ICD-10 codes.
Suppl. 2. Pyelonephritis ICD-10 codes.
Suppl. 3. Incidence of cystitis among emergency department all pediatric patients from 2016–2023.
Suppl. 4. Incidence of cystitis among emergency department pediatric patients aged <2 years from 2016–2023.
Suppl. 5. Incidence of cystitis among emergency department pediatric patients 2–17 years of age from 2016–2023.
Suppl. 6. Incidence of pyelonephritis among emergency depart-

ment all pediatric patients from 2016–2023.

Suppl. 7. Incidence of pyelonephritis among emergency department pediatric patients aged <2 years from 2016–2023.

Suppl. 8. Incidence of pyelonephritis among emergency department pediatric patients 2–17 years of age from 2016–2023.

Suppl. 9. Antibiotic medication classes administered among emergency department pediatric patients 0-17 years admitted with cystitis from 2016–2023.

Suppl. 10. Antibiotic medication classes administered among emergency department pediatric patients <2 years admitted with cystitis from 2016–2023.

Suppl. 11. Antibiotic medication classes administered among emergency department pediatric patients 2–17 years admitted with cystitis from 2016–2023.

Suppl. 12. Antibiotic medication classes administered among emergency department pediatric patients 0–17 years admitted with pyelonephritis from 2016–2023.

Suppl. 13. Antibiotic medication classes administered among emergency department pediatric patients <2 years admitted with pyelonephritis from 2016–2023.

Suppl. 14. Antibiotic medication classes administered among emergency department pediatric patients 2–17 years admitted with pyelonephritis from 2016–2023.

Suppl. 15. Antibiotic prescriptions among discharged emergency department pediatric patients 0–17 years with cystitis from 2016–2023.

Suppl. 16. Antibiotic prescriptions among discharged emergency department pediatric patients <2 years with cystitis from 2016–2023.

Suppl. 17. Antibiotic prescriptions among discharged emergency department pediatric patients 2–17 years with cystitis from 2016–2023.

Suppl. 18. Antibiotic prescriptions among discharged emergency department pediatric patients 2–17 years with pyelonephritis from 2016–2023.

Supplementary materials are available from https://doi.org/10.15441/ceem.25.055.

REFERENCES

1. Tullus K, Shaikh N. Urinary tract infections in children. Lancet 2020;395:1659–68.
2. Shaikh N, Morone NE, Bost JE, Farrell MH. Prevalence of urinary tract infection in childhood: a meta-analysis. Pediatr Infect Dis J 2008;27:302–8.
3. Copp HL, Shapiro DJ, Hersh AL. National ambulatory antibiotic prescribing patterns for pediatric urinary tract infection, 1998-2007. Pediatrics 2011;127:1027–33.
4. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med 2007;4:e297.
5. Subcommittee on Urinary Tract Infection, Steering Committee on Quality Improvement and Management; Roberts KB. Urinary tract infection: clinical practice guideline for the diagnosis and management of the initial UTI in febrile infants and children 2 to 24 months. Pediatrics 2011;128:595–610.
6. Fox MT, Amoah J, Hsu AJ, Herzke CA, Gerber JS, Tamma PD. Comparative effectiveness of antibiotic treatment duration in children with pyelonephritis. JAMA Netw Open 2020;3:e203951.
7. Miller AC, Arakkal AT, Sewell DK, et al. Comparison of different antibiotics and the risk for community-associated Clostridioides difficile infection: a case-control study. Open Forum Infect Dis 2023;10:ofad413.
8. Edlin RS, Shapiro DJ, Hersh AL, Copp HL. Antibiotic resistance patterns of outpatient pediatric urinary tract infections. J Urol 2013;190:222–7.
9. Gottlieb M, Wusterbarth E, Moyer E, Bernard K. Diverticulitis evaluation and management among United States emergency departments over an eight-year period. Am J Emerg Med 2024;86:83–6.
10. Faine BA, Rech MA, Vakkalanka P, et al. High prevalence of fluoroquinolone-resistant UTI among US emergency department patients diagnosed with urinary tract infection, 2018-2020. Acad Emerg Med 2022;29:1096–105.
11. Gaber CE, Kinlaw AC, Edwards JK, et al. Comparative effectiveness and harms of antibiotics for outpatient diverticulitis: two nationwide cohort studies. Ann Intern Med 2021;174:737–46.
12. Jackson MA, Schutze GE; Committee on Infectious Diseases. The use of systemic and topical fluoroquinolones. Pediatrics 2016;138:e20162706.
13. Mahony M, McMullan B, Brown J, Kennedy SE. Multidrug-resistant organisms in urinary tract infections in children. Pediatr Nephrol 2020;35:1563–73.

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):44-52
https://doi.org/10.15441/ceem.25.075

Original Article

Comparison of four noninvasive tools for predicting sepsis and septic shock mortality: a prospective cohort study

Matteo Guarino[1,2], Giacomo Maroncelli[1], Benedetta Perna[1], Paolo Baldin[1], Caterina Ghirardi[1,2], Alex Zanotto[1], Silvia Eichner[1], Matteo Bolognesi[1], Chiara Pesci[2], Martina Maritati[3], Carlo Contini[3], Roberto De Giorgio[1], Michele Domenico Spampinato[1,2]

[1]Department of Translational Medicine, University of Ferrara, Ferrara, Italy
[2]Emergency Department, S. Anna University Hospital of Ferrara, Ferrara, Italy
[3]Infectious Diseases Unit, Department of Medical Sciences, University of Ferrara, Ferrara, Italy

Objective Sepsis, a life-threatening organ dysfunction, is a major global health concern. Early detection remains challenging due to nonspecific symptoms. Noninvasive tools such as the shock index, diastolic shock index, capillary refill time (CRT), and mottling score (MS) could help clinicians assess hemodynamic status and predict mortality, but a comprehensive comparison of their prognostic value is lacking. This study compares the performance of those four tools in predicting mortality in septic patients at 24 hours, 7 days, and 28 days.

Methods A single-center, prospective observational study was conducted from January to September 2024. Adult patients (≥18 years) with suspected infection and a National Early Warning Score-2 of ≥5 were enrolled. Demographic data, vital signs, and CRT and MS results were collected at presentation, and mortality outcomes were recorded at 24 hours, 7 days, and 28 days.

Results In total, 135 patients were included (median age, 85 years [interquartile range, 79–90 years]; 44.4% female). The mortality rates were 15.6% at 24 hours, 25.2% at 7 days, and 35.6% at 28 days. CRT showed the highest predictive value for 24-hour mortality (area under the curve [AUC], 0.829; 95% confidence interval [CI], 0.755–0.889), and MS had the best performance at 7 days (AUC, 0.732; 95% CI, 0.646–0.806) and at 28 days (AUC, 0.749; 95% CI, 0.662–0.823). No significant differences emerged in pairwise comparisons.

Conclusion Although no one tool significantly outperformed the others, all four tools may provide useful, noninvasive mortality prediction in sepsis. CRT may be most effective for early risk stratification and MS correlates with mid-term outcomes, supporting their integration into clinical assessments.

Keywords Capillary refill time; Diastolic shock index; Mottling score; Sepsis; Septic shock

pISSN: 3022-1811
eISSN: 2383-4625

Received: April 9, 2025
Revised: August 7, 2025
Accepted: August 9, 2025

Correspondence to:
Roberto De Giorgio
(dgrrrt@unife.it)

How to cite this article:
Guarino M, Maroncelli G, Perna B, Baldin P, Ghirardi C, Zanotto A, Eichner S, Bolognesi M, Pesci C, Maritati M, Contini C, Giorgio RD, Spampinato MD. Comparison of four noninvasive tools for predicting sepsis and septic shock mortality: a prospective cohort study.
Clin Exp Emerg Med 2026;13(1):44-52.
https://doi.org/10.15441/ceem.25.075

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

Sepsis is a life-threatening organ dysfunction caused by a dysregulated host response to infection [1]. According to the Sepsis-3 consensus definition, sepsis occurs when an infection leads to acute changes in the Sequential Organ Failure Assessment (SOFA) score, indicating organ dysfunction [1]. Septic shock, a subset of sepsis, is characterized by persistent hypotension despite adequate fluid resuscitation and significantly increases the risk of in-hospital mortality [1,2]. The

© 2026 The Korean Society of Emergency Medicine

CEEM

Capsule Summary

What is already known

Noninvasive tools such as capillary refill time (CRT) and the mottling score (MS) are used to assess hemodynamic status in septic patients, but a comprehensive comparison of their predictive efficacy is lacking.

What is new in the current study

This study finds that CRT is the strongest predictor of early mortality (24 hours), and MS is more effective for predicting mortality from sepsis at 7 and 28 days. The findings support integrating CRT and MS into clinical protocols to enhance early and ongoing risk assessment for septic patients, potentially improving management strategies.

global incidence of sepsis is alarming, with estimates suggesting that it affects more than 30 million individuals annually, resulting in approximately 6 million deaths worldwide. In high-income countries, the overall mortality rate for sepsis ranges from 15% to 30%, whereas in low- and middle-income countries, it exceeds 50% [3,4]. This underscores the critical need for effective early detection and management strategies [5]. Early diagnosis and prognosis of sepsis remain significant challenges in clinical practice. The initial presentation of sepsis can be insidious, with nonspecific symptoms that can mimic less severe conditions. This often results in delayed recognition and treatment, so establishing protocols for early recognition and treatment is crucial for improving patient outcomes. Current guidelines emphasize the importance of early identification, with recommendations to assess patients for risk factors and clinical signs of sepsis within the first hours of presentation [2]. However, despite advances in diagnostic criteria and the introduction of early warning scores (e.g., National Early Warning Score-2 [NEWS-2]), timely identification is still insufficient, particularly in emergency settings. As a result, many patients continue to present with advanced disease, and sepsis morbidity and mortality rates remain high [2,5].

Given the importance of early assessment, noninvasive prognostic tools have emerged as valuable adjuncts in the management of septic patients. These tools, including the shock index (SI), diastolic SI (DSI), capillary refill time (CRT), and mottling score (MS), allow clinicians to rapidly evaluate a patient's hemodynamic status and perfusion [6–9]. The SI, defined as the ratio of the heart rate to systolic blood pressure, provides insight into a patient's cardiovascular status [6]. The DSI, an adaptation focusing on diastolic blood pressure, is intended to enhance the prognostic accuracy of the SI [7]. Both the SI and DSI have been proposed as potential early markers of circulatory impairment before overt hypotension occurs [6,7]; however, a comprehensive comparison of their prognostic efficacy is lacking. CRT assesses peripheral perfusion and has been shown to correlate with in-hospital mortality in critically ill patients [8]. The MS evaluates skin perfusion and can identify microcirculatory failure [9]. Each of these tools contributes to the early identification of patients at high risk for deterioration, allowing for timely intervention.

Despite the individual merits of these noninvasive assessments, no comprehensive studies have compared their efficacy in predicting sepsis mortality at various time frames. Therefore, our aim in this study was to evaluate and compare the prognostic performance of the SI, DSI, CRT, and MS in predicting 24-hour, 7-day, and 28-day mortality in septic patients. By conducting pairwise comparisons among these tools, we sought to identify which assessment method offers the greatest predictive value in the early stages of sepsis management. Ultimately, this research aims to enhance clinical decision-making and improve outcomes for patients experiencing sepsis and septic shock.

METHODS

Ethics statement

This study was approved by the Ethics Committee of the Area Vasta Emilia Centro (No. 301/2023/Oss/AOUFe). Written informed consent was obtained from all patients or their legally authorized representatives. The study was conducted in accordance with the ethical standards of the responsible committee on human experimentation and the 1975 Declaration of Helsinki.

Study design and setting

This single-center, observational, prospective, and noninterventional study was performed between January and September 2024. The study was conducted in a tertiary care center with approximately 600 beds and an emergency department that manages more than 70,000 visits annually. The inclusion criteria were as follows: (1) age ≥18 years; (2) suspicion of infectious disease;

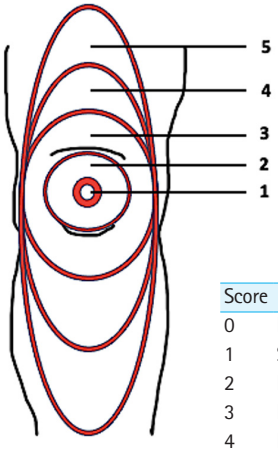
(3) NEWS-2 of ≥5; and (4) a signed informed consent form from each involved patient (or their relatives in cases with overall severe clinical conditions). Sepsis was confirmed using the Sepsis-3 diagnostic criteria [1]. NEWS-2 was selected as an inclusion criterion due to its widespread clinical use as an early indicator of sepsis risk [2,5]. The collected data were age, sex, vital parameters (i.e., systolic and diastolic blood pressure, heart rate, respiratory rate, peripheral oxygen saturation, oxygen requirement, body temperature, and neurological status), CRT, MS, and 24-hour, 7-day, and 28-day mortality from sepsis that were confirmed using medical records. All noninvasive parameters (SI, DSI, CRT, MS) were measured at the time of initial clinical evaluation in the emergency department, prior to any therapeutic intervention. CRT was measured by applying pressure to the distal phalanx for 5 seconds and assessing the time for color return [8]. The MS was evaluated using a validated five-grade scale (Fig. 1) [9], though interobserver reliability was not assessed in this study. It should be noted that the MS was assessed exclusively at the anterior aspect of the knee, according to the original validation study [9]. Although that site offers a standardized and practical approach, mottling is a systemic phenomenon that can manifest in other regions, such as the hands, feet, or abdomen. This limitation might reduce the sensitivity of the score in patients with atypical perfusion patterns or in those with dark skin, which can make mottling less visible. Future studies should consider broader anatomical assessments to improve diagnostic accuracy.

Statistical analysis

Categorical data are presented as absolute frequencies and percentages, and continuous variables are expressed as medians and interquartile ranges (IQRs). The Mann-Whitney test was used to compare continuous variables between two independent groups, and the chi-square test was used to compare the distribution of categorical data between two independent groups. Associations between two continuous variables were assessed using Spearman correlation. A multivariable logistic regression on 24-hour mortality was performed to adjust for potential confounding variables (age, Charlson Comorbidity Index [CCI], and altered mental status). A second multivariable logistic regression on the same outcomes was performed to adjust for SOFA scores and lactates. The discriminatory ability of each predictive score was evaluated using the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. Pairwise comparisons of ROC curves were conducted using the DeLong test.

To further explore the prognostic performance of the noninvasive tools under investigation, we conducted a series of additional statistical analyses. First, for each parameter (SI, DSI, CRT, and MS), we used the Youden index to determine the optimal cutoff values for predicting 24-hour, 7-day, and 28-day mortality. Based on those thresholds, we calculated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall accuracy. In addition, we performed a Kendall tau-b correlation analysis to assess the relationships among the four tools and evaluate whether they reflect overlapping or distinct physiological domains. Lastly, we conducted an age-stratified analysis by dividing the study population into four predefined age groups: <65, 65–75, 75–85, and >85 years. Notably, no deaths occurred in the <65 years group, preventing an outcome analysis in that category.

Statistical analyses were conducted using jamovi ver. 2.5 (the jamovi project) and MedCalc ver. 19.8 (MedCalc Software Ltd).



Score	Description
0	No mottling
1	Small mottling area around the knee
2	Mottling extending above the kneecap
3	Mottling reaching the mid-thigh
4	Mottling extending to the groin crease
5	Severe mottling spreading beyond the groin

Fig. 1. A graphical representation of the mottling score. As the severity increases, in-hospital mortality rises accordingly.

RESULTS

We prospectively enrolled 135 patients in this study, and their median age was 85 years (IQR, 79–90 years). Among them 60 (44.4%) were female. The mortality rate was 15.6% at 24 hours, 25.2% at 7 days, and 35.6% at 28 days. Further details are summarized in Table 1.

Among the tested tools, CRT was the best predictor of mortality at 24 hours, with an AUC of 0.829 (MS, 0.827; SI, 0.754; DSI, 0.755; P<0.001), indicating strong prognostic value in the early phase of sepsis (Fig. 2A). For longer-term mortality prediction, the MS emerged as the most reliable tool, with the best AUCs at both 7 days (MS, 0.732; SI, 0.701; DSI, 0.701; CRT, 0.667; P<0.001) (Fig. 2B) and 28 days (MS, 0.749; SI, 0.691; CRT, 0.687; DSI,

Table 1. Main clinical features and outcomes of enrolled patients

Characteristic	Total (n = 135)	24-hr mortality		7-day mortality		28-day mortality	
		Survivors (n = 114, 84.4%)	Nonsurvivors (n = 21, 15.6%)	Survivors (n = 101, 74.8%)	Nonsurvivors (n = 34, 25.2%)	Survivors (n = 87, 64.4%)	Nonsurvivors (n = 48, 35.6%)
Age (yr)	85 (79–90)	86 (79–91)	83 (80–89)	87 (79–91)	85 (79–90)	86 (79–90)	86 (78–90)
Female sex	60 (44.4)	49 (43.0)	11 (52.4)	39 (38.6)	21 (61.4)	31 (35.6)	29 (64.4)
CCI	3 (1–4)	3 (2–5)	2 (1–3)	3 (2–5)	3 (1–4)	3 (1–4)	3 (1–5)
SBP (mmHg)	100 (90–130)	105 (95–130)	85 (70–100)	110 (95–135)	90 (70–110)	110 (95–135)	90 (75–110)
DBP (mmHg)	60 (50–80)	60 (60–80)	50 (40–60)	60 (60–80)	50 (40–60)	60 (60–80)	60 (40–70)
HR (bpm)	102 (90–115)	100 (90–113)	105 (90–125)	101 (90–112)	105 (90–120)	100 (90–110)	105 (90–120)
RR (breaths/min)	24 (20–28)	24 (20–28)	28 (22–32)	24 (20–28)	24 (20–28)	24 (18–28)	27 (22–30)
SpO_2 (%)	93 (90–96)	94 (91–96)	92 (90–96)	94 (91–97)	93 (90–96)	94 (91–97)	93 (90–96)
FiO_2 (%)	0.24 (0.21–0.40)	0.21 (0.21–0.35)	0.40 (0.21–0.80)	0.21 (0.21–0.36)	0.29 (0.21–0.60)	0.21 (0.21–0.35)	0.28 (0.21–0.60)
GCS score	14 (11–15)	14 (12–15)	9 (8–12)	14 (12–15)	10 (8–13)	14 (13–15)	11 (9–14)
Body temperature (°C)	38.0 (36.9–38.5)	38.0 (36.9–38.5)	37.8 (37.0–38.2)	37.9 (36.9–38.4)	38.0 (37.0–38.4)	38.0 (36.9–38.3)	37.8 (36.7–38.3)
DSI	1.5 (1.3–1.9)	1.5 (1.3–1.8)	2.0 (1.8–2.9)	1.5 (1.3–1.8)	2.0 (1.4–2.6)	1.5 (1.3–1.8)	1.8 (1.4–2.5)
SI	0.9 (0.7–1.1)	0.9 (0.7–1.1)	1.2 (1.0–1.9)	0.9 (0.7–1.1)	1.1 (0.9–1.6)	0.9 (0.7–1.1)	1.1 (0.9–1.5)
CRT (sec)	3.5 (3.0–5.0)	3.0 (2.5–5.0)	6.0 (5.0–8.0)	3.3 (2.5–5.0)	5.0 (3.0–7.0)	3.0 (2.0–4.5)	5.0 (3.0–7.0)
MS	1 (1–2)	1 (0–2)	3 (2–4)	1 (0–2)	3 (1–4)	1 (0–2)	2 (1–3)
SOFA score	5 (4–7)	5 (3–6)	8 (6–9)	4 (3–6)	8 (6–9)	4 (3–6)	7 (5–9)

Values are presented as median (interquartile range) or number (%).

CCI, Charlson Comorbidity Index; SBP, systolic blood pressure; DBP, diastolic blood pressure; HR, heart rate; bpm, beats per minute; RR, respiratory rate; SpO_2, peripheral oxygen saturation; FiO_2, fraction of inspired oxygen; GCS, Glasgow Coma Scale; DSI, diastolic shock index; SI, shock index; CRT, capillary refill time; MS, mottling score; SOFA, Sequential Organ Failure Assessment.

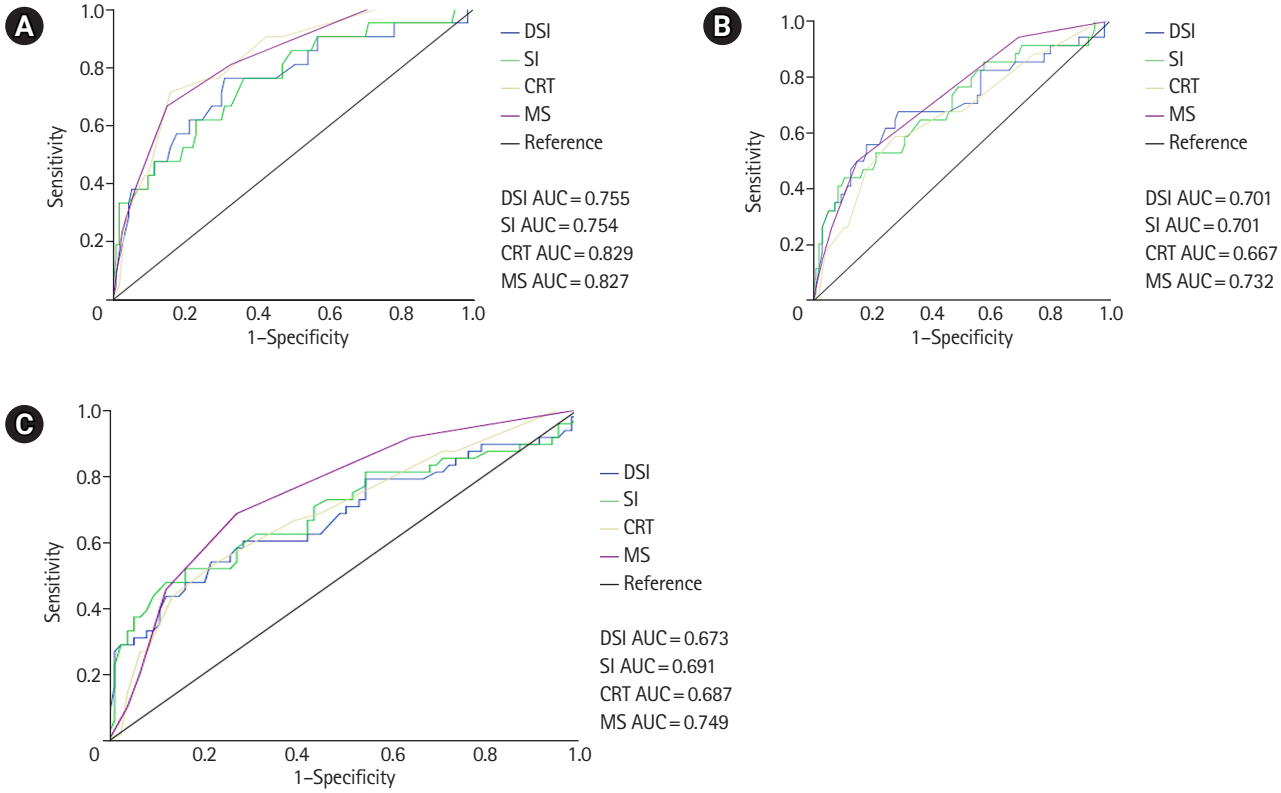


Fig. 2. Receiver operating characteristic curves for diastolic shock index (DSI), SI, capillary refill time (CRT), and mottling score (MS) in assessing (A) 24-hour mortality, (B) 7-day mortality, and (C) 28-day mortality from sepsis and septic shock. AUC, area under the curve.

0.673; P<0.001) (Fig. 2C). No statistically significant differences were observed among the four tools in the pairwise comparisons (Table 2). In the univariable analysis, all four tools were found to be independent risk factors for mortality at 24 hours, 7 days, and 28 days (Table 3). These results were further explored through a corrected multivariable logistic regression analysis (Table 4), which evaluated each noninvasive tool (SI, DSI, CRT, and MS) independently while adjusting for age, CCI, and altered mental status. All four tools were found to be independently associated with mortality across all time points. Specifically, SI showed an odds ratio (OR) of 10.949 (95% confidence interval [CI], 4.065–29.480) for 24-hour mortality, 7.299 (95% CI, 3.012–17.689) for 7-day mortality, and 10.379 (95% CI, 4.070–26.467) for 28-day mortality. DSI had ORs of 4.492 (95% CI, 2.402–8.638), 3.960 (95% CI, 2.225–7.045), and 4.116 (95% CI, 2.317–7.312), respectively, at the same time points. CRT also retained statistical significance in all models with ORs of 1.810 (95% CI, 1.451–2.258), 1.422 (95% CI, 1.195–1.692), and 1.358 (95% CI, 1.149–1.605), respectively. Similarly, MS was associated with increased mortality with ORs of 3.007 (95% CI, 2.117–4.270), 2.203 (95% CI, 1.679–2.890), and 2.103 (95% CI, 1.609–2.748) at 24 hours, 7 days, and 28 days, respectively. In addition, we used the Youden index to identify optimal cutoff values for each tool and calculated corresponding sensitivity, specificity, PPV, NPV, and accuracy metrics. Those results are presented in Suppl. 1 and provide clinically useful thresholds to guide bedside interpretation of results from these tools.

To assess whether the tools reflect overlapping physiological domains or offer complementary information, we performed a Kendall tau-b correlation analysis. SI and DSI were closely associated with parameters such as PaO_2/FiO_2 (fraction of inspired oxygen), platelet count, and hepatic dysfunction. In contrast, CRT and MS showed strong correlations with mean arterial pressure and the Glasgow Coma Scale. Full results are shown in Suppl. 2.

In addition to the primary analyses, a supplementary multivariable analysis was conducted for each noninvasive parameter independently, controlling for total SOFA scores and lactate levels.

Table 3. Univariable analysis correlating each of the tested tools with 24-hour, 7-day, and 28-day mortality

Variable	OR (95% CI)	P-value
24-hr mortality		
DSI	4.911 (2.182–11.056)	<0.001
SI	13.657 (3.869–48.211)	<0.001
CRT (sec)	1.896 (1.435–2.504)	<0.001
MS	2.371 (1.643–3.422)	<0.001
7-day mortality		
DSI	3.949 (1.907–8.177)	<0.001
SI	8.549 (2.766–26.428)	<0.001
CRT (sec)	1.375 (1.120–1.688)	0.002
MS	1.817 (1.353–2.440)	<0.001
28-day mortality		
DSI	3.638 (1.769–7.481)	<0.001
SI	8.911 (2.790–28.458)	<0.001
CRT (sec)	1.438 (1.171–1.764)	0.001
MS	1.889 (1.399–2.553)	<0.001

OR, odds ratio; CI, confidence interval; DSI, diastolic shock index; SI, shock index; CRT, capillary refill time; MS, mottling score.

Table 2. AUC comparison among DSI, SI, CRT, and MS in terms of 24-hour, 7-day, and 28-day mortality and pairwise comparisons of the ROC curves.

Variable	AUC	95% CI	P-value	Pairwise comparison of ROC curve P-value			
				DSI	SI	CRT	MS
24-hr mortality			<0.001				
DSI	0.755	0.672–0.824		-	0.948	0.168	0.259
SI	0.754	0.674–0.825		0.948	-	0.132	0.244
CRT (sec)	0.829	0.755–0.889		0.168	0.132	-	0.970
MS	0.827	0.753–0.887		0.259	0.244	0.970	-
7-day mortality			<0.001				
DSI	0.701	0.614–0.779		-	>0.999	0.622	0.592
SI	0.701	0.614–0.779		>0.999	-	0.606	0.579
CRT (sec)	0.667	0.579–0.748		0.622	0.606	-	0.174
MS	0.732	0.646–0.806		0.592	0.579	0.174	-
28-day mortality			<0.001				
DSI	0.673	0.582–0.756		-	0.488	0.818	0.180
SI	0.691	0.601–0.772		0.488	-	0.955	0.317
CRT (sec)	0.687	0.597–0.769		0.818	0.955	-	0.140
MS	0.749	0.662–0.823		0.180	0.317	0.140	-

The P-values refer to pairwise comparisons of the ROC curves for the different tools (DSI, SI, CRT, and MS) in predicting 24-hour, 7-day, and 28-day mortality. These P-values indicate whether these tools differ with statistical significance in their predictive performance.
AUC, area under the curve; DSI, diastolic shock index; SI, shock index; CRT, capillary refill time; MS, mottling score; ROC, receiver operating characteristic; CI, confidence interval.

Table 4. Multivariable logistic regression models evaluating the associations between each noninvasive tool and mortality at 24 hours, 7 days, and 28 days

Model	24-hr mortality		7-day mortality		28-day mortality	
	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value
Shock index	10.949 (4.065–29.480)	<0.001	7.299 (3.012–17.689)	<0.001	10.379 (4.070–26.467)	<0.001
Age (yr)	1.005 (0.970–1.042)	0.776	1.003 (0.972–1.036)	0.848	1.012 (0.981–1.043)	0.449
CCI	0.990 (0.853–1.148)	0.890	0.992 (0.868–1.113)	0.901	1.135 (0.990–1.301)	0.070
Altered mentation	17.914 (2.283–140.564)	0.006	6.881 (2.240–21.139)	0.001	3.662 (1.596–8.400)	0.002
Diastolic shock index	4.492 (2.402–8.638)	<0.001	3.960 (2.225–7.045)	<0.001	4.116 (2.317–7.312)	<0.001
Age (yr)	1.005 (0.969–1.042)	0.799	1.002 (0.970–1.035)	0.901	1.010 (0.980–1.042)	0.509
CCI	0.993 (0.856–1.152)	0.928	0.992 (0.868–1.113)	0.901	1.122 (0.982–1.283)	0.092
Altered mentation	18.214 (2.341–141.698)	0.006	7.248 (2.342–22.425)	0.001	3.820 (1.659–8.795)	0.002
Capillary refill time (sec)	1.810 (1.451–2.258)	<0.001	1.422 (1.195–1.692)	<0.001	1.358 (1.149–1.605)	<0.001
Age (yr)	1.002 (0.964–1.042)	0.904	0.999 (0.968–1.032)	0.966	1.006 (0.977–1.036)	0.669
CCI	0.923 (0.787–1.080)	0.321	0.957 (0.836–1.095)	0.523	1.096 (0.961–1.249)	0.172
Altered mentation	9.979 (1.277–77.962)	0.028	4.899 (1.613–14.882)	0.005	2.720 (1.214–6.095)	0.015
Mottling score	3.007 (2.117–4.270)	<0.001	2.203 (1.679–2.890)	<0.001	2.103 (1.609–2.748)	<0.001
Age (yr)	0.998 (0.958–1.040)	0.925	0.996 (0.963–1.029)	0.805	1.003 (0.973–1.034)	0.856
CCI	0.936 (0.789–1.111)	0.450	0.964 (0.835–1.114)	0.622	1.122 (0.976–1.291)	0.106
Altered mentation	21.039 (2.199–201.259)	0.008	6.040 (1.847–19.752)	0.003	2.960 (1.259–6.957)	0.013

Each model was stratified by age, CCI, and altered mentation.
OR, odds ratio; CI, confidence interval; CCI, Charlson Comorbidity Index.

CRT and MS emerged as independent predictors of 24-hour mortality, even when adjusted for established prognostic markers. Conversely, SI and DSI did not demonstrate independent predictive value within these adjusted models, suggesting that their prognostic utility might be contingent on concurrent hemodynamic data (Suppl. 3).

Finally, we conducted an age-stratified ROC analysis to evaluate prognostic performance across different age groups. Patients were divided into four predefined age groups (<65, 65–75, 75–85, and >85 years); however, due to the absence of deaths in the <65 years group, the ROC analysis was limited to the remaining three groups. The curves presented in Suppl. 4 confirm consistent trends across strata, with CRT performing better in early mortality prediction and MS showing stronger performance at the later time points.

DISCUSSION

The results of this study show that CRT had the highest AUC for predicting mortality at 24 hours, whereas MS had the highest AUC for predicting mortality at 7 and 28 days. Furthermore, pairwise comparisons among the tools did not reveal significant differences, suggesting that each tool has merit, and they can be used interchangeably in practice without one clearly outperforming the others.

However, it is important to acknowledge that the tools evaluated in this study reflect distinct physiological domains: SI and DSI are primarily indicators of macrocirculatory function [6,7], whereas CRT and MS assess microcirculatory status [8,9]. Our intention was to compare these tools side by side to evaluate their relative prognostic performance, despite their differing physiological targets. This approach allows clinicians to understand whether one domain offers superior predictive value in practice. Nevertheless, we recognize that future studies could benefit from grouping these tools by physiological domain or exploring composite models that integrate both macro- and microcirculatory indicators to enhance risk stratification.

The strong predictive value of CRT at 24 hours underscores its utility as a rapid assessment tool in the emergency setting. CRT is a simple bedside test that evaluates peripheral perfusion and can be performed quickly, making it an accessible option for frontline clinicians [8,10,11]. In septic patients, abnormal CRT, which reflects impaired peripheral perfusion, has been associated with higher in-hospital mortality and worse outcomes [8,10]. Research shows that a prolonged CRT after fluid resuscitation correlates with a high risk of adverse events, including longer intensive care unit stays, need for mechanical ventilation, and renal replacement therapy [10]. One meta-analysis of multiple studies demonstrated that prolonged CRT is a reliable predictor of in-hospital mortality in septic patients, particularly those with acute circulatory failure [11]. The significant correlation between CRT and in-hospital mortality in our study highlights the importance of hemodynamic status early in the course of sepsis. Delayed or inadequate perfusion is a critical indicator of worsening clinical

condition [10], and CRT serves as a timely alert for healthcare providers.

In contrast, the superiority of MS in predicting 7- and 28-day mortality suggests its effectiveness in assessing mid-term outcomes in septic patients. This indicates that although immediate interventions based on CRT are crucial, ongoing assessment with the MS can provide valuable insights into a patient's evolving clinical picture. Mottling, a manifestation of microcirculatory failure, reflects the body's systemic response to sepsis and can be an important indicator of severe disease progression [9,12–14]. The ability of MS to remain a reliable predictor over time can be attributed to its assessment of skin perfusion and overall hemodynamic stability. Studies have demonstrated that a higher MS is associated with worse outcomes in septic shock [12]. For instance, one study showed that patients with a higher MS had significantly lower skin oxygen saturation and higher mortality at 28 days [12]. The MS, which ranges from 0 (no mottling) to 5 (severe mottling), can be a quick bedside indicator of the severity of septic shock and help to predict in-hospital mortality. Patients with mottling scores of 3 or more typically have a poor prognosis [12,13].

Moreover, the duration of mottling is also a critical factor. Persistent mottling for more than 6 hours is linked with increased in-hospital mortality, emphasizing the importance of early detection and management of circulatory failure in septic shock patients [14]. However, the MS has limitations, particularly in darker-skinned individuals, and requires training for consistent interpretation [12–14].

Both SI and DSI generally correlate with the development of hypotension. Nonetheless, although hypotension is indeed a late feature of shock states, both SI and DSI have been used as potential markers of early sepsis due to their ability to detect circulatory impairment before overt hypotension occurs [6,7]. These indexes provide valuable insights into a patient's hemodynamic status, allowing for earlier identification and intervention in septic patients, which is crucial for improving outcomes.

The median age of our study population (85 years) can affect the interpretation of our results. Older patients typically exhibit a more complex clinical picture, including multiple comorbidities that can influence both sepsis presentation and outcomes [5,15,16]. The frailty and diminished physiological performance common in elderly people can also contribute to an increased risk of in-hospital mortality [15]. Other factors that might affect elderly patients' prognosis are age-related organ decline and delayed diagnosis due to atypical symptom presentation [17]. Our findings highlight the need for tailored assessment strategies in these patients, in whom traditional metrics might not fully capture the nuances of sepsis presentation and progression. Additionally, the small sample size raises concerns about the generalizability of our findings. Although our results provide valuable insights, larger studies are needed to validate these findings and enhance their statistical power. The lack of significant differences in the pairwise comparisons indicates that both CRT and the MS are promising tools, but further research with a more substantial cohort could reveal nuanced differences in their predictive capabilities that were not evident in our study.

The clinical implications of our findings are significant. The early identification of patients at high risk of in-hospital mortality is critical to optimize treatment strategies in sepsis. Although we did not find substantial differences in the predictive capabilities of the assessed tools, the strong performance of CRT and the MS suggests that they could serve as complementary assessments. For instance, CRT could be used as an initial rapid assessment, followed by the MS to monitor ongoing risk over time. Incorporating these noninvasive tools into routine clinical practice could help clinicians recognize deteriorating patients and prompt timely interventions. Moreover, their ease of use makes them particularly valuable in resource-limited settings, especially where access to advanced monitoring technologies is restricted.

In the multivariable analysis, each tool demonstrated a significant association with mortality at all three tested time points, indicating that they are valuable predictors of patient outcomes in sepsis. Despite their differences in focus (from cardiovascular function to peripheral perfusion), all four tools consistently identified patients at a high risk of death at specific time points. This result confirms that, in clinical practice, any of these scores can be used effectively to stratify risk in septic patients, supporting timely and targeted interventions to improve survival rates.

Our findings are further supported by additional analyses. The multivariable logistic regression models confirm that each tool independently predicts mortality at all time points, even after adjusting for key confounders such as age, comorbidity burden, and mental status. This reinforces the robustness of their prognostic value in the clinical setting. We also identified optimal cutoff values for each parameter to provide practical thresholds for frontline application and decision-making. The correlation analysis showed moderate interdependence between tools, particularly within the same physiological domain, but no complete overlap, suggesting that combining macrocirculatory (SI, DSI) and microcirculatory (CRT, MS) markers could offer additive prognostic value. A subgroup analysis across elderly age groups confirmed the reproducibility of these trends in even very old populations, rein-

forcing the potential utility of these tools in geriatric sepsis assessment.

A strength of this study is its exclusive focus on noninvasive tools that require no laboratory testing or specialized equipment. Unlike biomarkers such as lactate, which require blood sampling and laboratory infrastructure, the tools evaluated here can be applied in any clinical setting (including prehospital care) at no cost. This makes them particularly valuable in resource-limited environments, where rapid risk stratification is essential but access to advanced diagnostics might be restricted.

Several limitations of this study warrant discussion. First, the study was conducted in a single center, limiting the generalizability of the findings. Patient populations, healthcare resources, and clinical practices vary significantly across hospitals, especially between countries with differing healthcare systems. Second, our small sample size could affect the statistical power of our results and our ability to detect subtle differences among the prognostic tools. Furthermore, a formal power calculation was not performed. A larger cohort would provide more robust and generalizable results. Finally, the median age of our study population was 85 years. Although that reflects the higher vulnerability of elderly patients to sepsis, it limits the applicability of the findings to younger or middle-aged patients, who can exhibit different physiological responses and outcomes. External validation in more diverse and age-balanced cohorts is warranted to confirm the applicability of our results across broader populations.

Future studies should aim to include larger and more diverse patient populations to further validate our findings. Additionally, our study did not include a control group, which limits some comparative analyses; however, our primary objective was to assess the predictive capacity of these tools in septic patients. Exploring the integration of these tools into clinical decision-making frameworks could provide insights into how they might enhance patient outcomes. Investigating potential combinations of these tools or their use in conjunction with traditional scoring systems (such as the SOFA score) could also yield promising results.

In conclusion, this study highlights the prognostic value of noninvasive tools in the assessment of sepsis mortality. CRT was the parameter with the highest AUC at 24 hours, whereas MS had the highest AUC at 7 and 28 days. Although our findings did not reveal significant differences among the tools, the complementary nature of these scores provides an opportunity to improve clinical practice. Furthermore, because each parameter independently predicted mortality at all three of the tested time points, future studies should explore whether combining the tools (potentially integrating both macro- and microcirculatory indicators) could enhance prognostic accuracy. These findings further support the hypothesis that microcirculatory markers, such as CRT and MS, retain independent prognostic relevance in the early phase of sepsis, even when accounting for comprehensive tools such as the SOFA score and serum lactate. The lack of predictive independence for SI and DSI in this context could reflect their stronger overlap with hemodynamic parameters already embedded in the SOFA score, raising the possibility of multicollinearity in multivariable models. Future studies should explore hybrid assessment strategies that integrate both macrocirculatory and microcirculatory indices to optimize early risk stratification. Further research is needed to enhance our understanding of the roles of these tools in the management of sepsis, particularly in vulnerable populations such as elderly people.

ARTICLE INFORMATION

Author contributions

Conceptualization: MG; Data curation: MG, GM, PB, CG, AZ, SE, MB, MDS; Formal analysis: MG, BP, MDS; Validation: MG, BP, CP, MM, CC, RdeG; Writing–original draft: MG; Writing–review & editing: all authors. All authors read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

Funding

The authors received no financial support for this study.

Data availability

Data analyzed in this study are not publicly available due to privacy policy, but they are available from the corresponding author upon reasonable request.

Supplementary materials

Suppl. 1. Optimal cutoff values and diagnostic performance metrics for each noninvasive tool (SI, DSI, CRT, MS) in predicting 24-hour, 7-day, and 28-day mortality.

Suppl. 2. Spearman correlation coefficients between each noninvasive prognostic tool (SI, DSI, CRT, and MS) and the SOFA score.

Suppl. 3. Multivariable logistic regression models evaluating the associations between SI, DSI, CRT, and MS and mortality at 24 hours, 7 days, and 28 days. Each model was stratified by SOFA scores and lactate levels.

Suppl. 4. Receiver operating characteristic (ROC) curves of SI, DSI, CRT, and MS for predicting 24-hour, 7-day, and 28-day mortality across four predefined age groups: <65, 65–75, 75–85, and >85 years.

Supplementary materials are available from https://doi.org/10.15441/ceem.25.075.

REFERENCES

1. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016;315:801–10.
2. Evans L, Rhodes A, Alhazzani W, et al. Surviving Sepsis Campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med 2021;47:1181–247.
3. Chiu C, Legrand M. Epidemiology of sepsis and septic shock. Curr Opin Anaesthesiol 2021;34:71–6.
4. World Health Organization (WHO). Global report on the epidemiology and burden of sepsis: current evidence, identifying gaps and future directions. WHO; 2020.
5. Guarino M, Perna B, Cesaro AE, et al. 2023 Update on sepsis and septic shock in adult patients: management in the emergency department. J Clin Med 2023;12:3188.
6. Diaztagle Fernández JJ, Castañeda-González JP, Trujillo Zambrano JI, Duarte Martínez FE, Saavedra Ortiz MÁ. Assessment of the shock index in septic shock: a systematic review. Med Intensiva (Engl Ed) 2024;48:e10–9.
7. Jeon Y, Kim S, Ahn S, et al. Predicting septic shock in patients with sepsis at emergency department triage using systolic and diastolic shock index. Am J Emerg Med 2024;78:196–201.
8. Kattan E, Bakker J, Estenssoro E, et al. Hemodynamic phenotype-based, capillary refill time-targeted resuscitation in early septic shock: the ANDROMEDA-SHOCK-2 randomized clinical trial study protocol. Rev Bras Ter Intensiva 2022;34:96–106.
9. Preda G, Bourcier S, Joffre J, et al. Mottling score is associated with 28-day mortality in critically ill patients with sepsis. Minerva Anestesiol 2017;83:664–6.
10. Castro R, Kattan E, Ferri G, et al. Effects of capillary refill time- vs. lactate-targeted fluid resuscitation on regional, microcirculatory and hypoxia-related perfusion parameters in septic shock: a randomized controlled trial. Ann Intensive Care 2020;10:150.
11. Jacquet-Lagrèze M, Pernollet A, Kattan E, et al. Prognostic value of capillary refill time in adult patients: a systematic review with meta-analysis. Crit Care 2023;27:473.
12. Kazune S, Caica A, Volceka K, Suba O, Rubins U, Grabovskis A. Relationship of mottling score, skin microcirculatory perfusion indices and biomarkers of endothelial dysfunction in patients with septic shock: an observational study. Crit Care 2019;23:311.
13. Ferraris A, Bouisse C, Mottard N, et al. Mottling score and skin temperature in septic shock: relation and impact on prognosis in ICU. PLoS One 2018;13:e0202329.
14. de Moura EB, Amorim FF, da Cruz Santana AN, et al. Skin mottling score as a predictor of 28-day mortality in patients with septic shock. Intensive Care Med 2016;42:479–80.
15. Ibarz M, Haas LE, Ceccato A, Artigas A. The critically ill older patient with sepsis: a narrative review. Ann Intensive Care 2024;14:6.
16. Stenberg H, Li X, Pello-Esso W, et al. The effects of sociodemographic factors and comorbidities on sepsis: a nationwide Swedish cohort study. Prev Med Rep 2023;35:102326.
17. Bouza C, Martínez-Alés G, López-Cuadrado T. The impact of dementia on hospital outcomes for elderly patients with sepsis: a population-based study. PLoS One 2019;14:e0212196.

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):53-64
https://doi.org/10.15441/ceem.25.153

Original Article

pISSN: 3022-1811
eISSN: 2383-4625

Automated chain-of-thought evaluation framework for large language model–generated emergency department documentation: a simulation-based study

Dasol Choi[1,2,3*], Junhyuk Seo[4,5,6*], Won Cul Cha[5,7], Minha Kim[5,7], Sejin Heo[5,7], Hansol Chang[7], Taerim Kim[6,7,8]

[1]Yonsei University, Seoul, Korea
[2]MODULABS, Seoul, Korea
[3]Now with AIM Intelligence, Seoul, Korea
[4]Department of Nursing, Samsung Medical Center, Seoul, Korea
[5]Department of Digital Health, Samsung Advanced Institute of Health Sciences and Technology (SAIHST), Sungkyunkwan University, Seoul, Korea
[6]Now with ETOILE Inc, Seoul, Korea
[7]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea
[8]ETOILE Inc, Seoul, Korea

Objective This study aimed to develop and validate MEDIVAL (Medical Documentation Validation), a progressive chain-of-thought (CoT) evaluation framework for automated assessment of large language model (LLM)-generated emergency department documentation, designed to align with expert clinical judgment in acute care settings.

Methods We designed a three-tier evaluation framework incorporating persona-based, error-enhanced, and insight-integrated strategies. The framework was tested across four LLMs (GPT-4o, GPT-4.1, Claude-3.5, Claude-3.7) on 33 emergency department records reviewed by four expert emergency physicians. Each model applied the three CoT strategies across five criteria: appropriateness, accuracy, structure/format, conciseness, and clinical validity. Model outputs were compared with expert ratings using Spearman correlation coefficients. Differences were analyzed with the Friedman test and Wilcoxon signed rank test with Bonferroni correction. Reproducibility was assessed through intraclass correlation coefficient (ICC) analysis.

Results All models demonstrated stronger alignment with expert ratings as CoT complexity increased, with Claude-3.7 (r=0.712, P<0.001) and GPT-4o (r=0.702, P<0.001) showing the highest correlations under the insight-integrated strategy. GPT-4.1 showed the greatest relative improvement (43.3% increase, r=0.457 to r=0.655, P<0.001). Significant overall differences were observed across strategies ($\chi^2(2)=48.39$, P<0.001), though the error-enhanced and insight-integrated approaches differed only modestly yet significantly (P=0.002). High reproducibility was confirmed (ICC >0.919), with Claude-3.5 achieving the most consistent results (ICC, 0.997–0.998).

Conclusion MEDIVAL demonstrates that progressive CoT strategies systematically improve automated evaluation of emergency department documentation while maintaining excellent reproducibility. This framework offers a viable prescreening tool to reduce expert workload and support reliable artificial intelligence integration into emergency medicine workflows.

Keywords Artificial intelligence; Medical documentation; Emergency department; Large language models; Clinical evaluation

Received: July 15, 2025
Revised: August 25, 2025
Accepted: August 31, 2025

Correspondence to:
Taerim Kim
(taerim.j.kim@gmail.com)

*Dasol Choi and Junhyuk Seo contributed equally to this study as co-first authors.

How to cite this article:
Choi D, Seo J, Cha WC, Kim M, Heo S, Chang H, Kim T. Automated chain-of-thought evaluation framework for large language model–generated emergency department documentation: a simulation-based study. Clin Exp Emerg Med 2026;13(1):53-64. https://doi.org/10.15441/ceem.25.153

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

© 2026 The Korean Society of Emergency Medicine

Capsule Summary

What is already known

Evaluating large language model (LLM)-generated emergency department documentation requires substantial expert time and resources, creating a major bottleneck for clinical adoption of automated systems in acute care.

What is new in the current study

A progressive chain-of-thought evaluation framework provides reliable alignment with expert assessment while maintaining high reproducibility across diverse LLM architectures, offering a practical prescreening solution for emergency medicine documentation workflows.

INTRODUCTION

Medical documentation is a cornerstone of healthcare delivery [1], enabling continuity of care and facilitating multidisciplinary communication [2–4]. However, it imposes a significant burden on physicians and nurses, often requiring additional hours to complete [5]. This documentation burden is now recognized as a major contributor to burnout [6–11], raising significant concerns about care quality and patient safety [12]. Accordingly, efforts to mitigate this burden, including optimization of electronic health records (EHRs) and workflow streamlining, have intensified [13–15]. Within this context, large language models (LLMs) have emerged as potential game-changers in healthcare through their ability to generate medical documentation [16–19]. While they have shown promise in generation, summarization, and standardization tasks, concerns remain regarding their accuracy and reliability [20].

However, evaluating LLM-generated output presents particular challenges. It requires extensive expert involvement and time-consuming processes, which can delay implementation in clinical settings [21,22]. Recent advances, however, suggest a shift in this paradigm. Increasing evidence shows that LLMs can serve not only as content generators but also as evaluators, effectively functioning as "LLM-as-a-judge" systems that may help reduce reliance on expert review [23,24]. Despite these developments, the effectiveness of different model architectures and prompting strategies for medical document evaluation remains underexplored.

Our previous study examined the feasibility of LLM-based medical documentation generation and proposed an evaluation framework combining qualitative and quantitative elements [21]. Through this dual approach, we identified recurring themes in expert feedback, particularly their emphasis on clinical accuracy, logical flow, and adherence to documentation standards. Experts consistently prioritized these elements when reviewing medical records, offering a basis for structuring LLM-based evaluation methods.

Building upon these insights, we developed an LLM evaluation framework with three progressive strategies—persona-based, error-enhanced, and insight-integrated—each reflecting distinct expert evaluation patterns. The present study applies this framework across several state-of-the-art LLMs (GPT-4o, GPT-4.1, Claude-3.5, and Claude-3.7) to examine their effectiveness in evaluating medical documentation. By benchmarking model performance against expert assessments, we evaluate their potential to reduce workload without compromising quality, addressing a critical gap in efficient evaluation approaches for clinical documentation.

This comparative analysis served two main purposes. First, it sought to validate the effectiveness of progressive chain-of-thought (CoT) prompting across multiple LLM architectures. Second, it examined the feasibility of LLM-based systems as prescreening tools for clinical documentation. We hypothesize that although baseline performance varies across models, applying structured reasoning via progressive CoT prompting consistently improves evaluation quality. A preliminary version of this work was presented as a short paper at MEDINFO 2025 [25]. This article extends that work through additional cross-model validation, comparative analyses, and expanded discussion.

METHODS

Ethics statement

This study used synthetic emergency department records and did not involve human subjects. The original data collection was approved by the Institutional Review Board of Samsung Medical Center (No. 2023-12-018-001).

Dataset and expert evaluation score baseline

We used the full set of 33 LLM-generated emergency department (ED) records and their corresponding expert evaluation scores from our previous study, which originated from a medical prompt challenge [21]. To generate these ED records, participants were given detailed virtual patient scenarios developed and validated by an ED professor and a registered nurse. Based on these scenarios, participants created prompts that were processed by the HyperCLOVA X (HCX-002; Naver Cloud) LLM model [26] to produce medical documentation.

Each ED record was required to follow a standardized format, as outlined in Table 1. To construct effective prompts, participants were provided with patient triage records, initial consultation transcripts, and physical examination data, all in Korean. The goal was to ensure that generated records adhered closely to the expected structure and contained comprehensive clinical details.

In the prior study, four medical experts, each with more than 10 years of clinical experience, evaluated every document using five criteria shown in Table 2 [21]. Scores were assigned on a 5-point Likert scale (1 [poor] to 5 [excellent]). Reliability of this evaluation process was thoroughly validated. Test-retest analysis showed high correlations ($r=0.776$, $P<0.001$), and interrater reliability demonstrated strong consistency (intraclass correlation coefficient [ICC], 0.653–0.887; $P<0.001$). Clinical validity achieved the highest agreement (ICC, 0.887), underscoring the consistency of expert judgment on this critical dimension. These five criteria now serve as the foundation for our current comparative analysis of different LLM models and CoT strategies in medical document evaluation.

Table 1. Standardized structure of emergency department records

Patient information	Clinical assessment	Medical decision-making
Chief complaint	Physical examination findings	Problem list
Vital signs	Review of systems	Differential diagnosis
History of present illness, past medical history	Personal and social history	Diagnostic plan, treatment plan

MEDIVAL: three-module CoT framework

We propose a CoT evaluation framework called MEDIVAL (Medical Documentation Validation) that systematically evaluates medical records through three distinct approaches (Fig. 1). CoT prompting is a technique that guides LLMs to decompose complex tasks into intermediate steps, simulating human reasoning processes [27]. In this framework, "modules" refer to prompting components with specific evaluation functions, whereas "strategies" represent progressive combinations of these modules into increasingly comprehensive evaluation approaches. Each strategy builds upon the prior one by incorporating additional modules, thereby producing a progressively more sophisticated evaluation process.

Table 2. Expert evaluation criteria for large language model–generated medical documents[a]

Criterion	Description	Focus area
Appropriateness	Adherence to medical documentation standards	Content relevance, proper medical terminology, professional tone
Accuracy	Correctness of medical information	Factual correctness, absence of hallucinations, completeness
Structure/format	Logical organization and formatting	Section organization, information flow, hierarchical presentation
Conciseness	Information density and brevity	Elimination of redundancy while maintaining completeness
Clinical validity	Practical usability in clinical context	Sound clinical reasoning, appropriate differential diagnosis, actionable treatment plans

[a]Developed in our previous study [21].

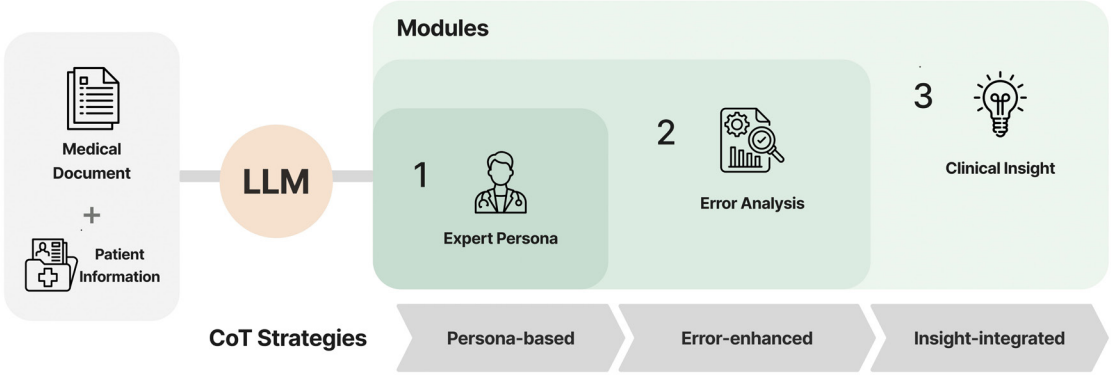


Fig. 1. Overview of the three-module chain-of-thought (CoT) evaluation framework for large language model (LLM)-based assessment. The framework integrates expert persona, error analysis, and clinical insight modules to progressively enhance evaluation quality.

Framework design and implementation

Our evaluation framework implemented three progressive evaluation strategies using multiple state-of-the-art LLMs: GPT-4o (OpenAI) [28], GPT-4.1 (OpenAI) [29], Claude-3.5 (Anthropic) [30], and Claude-3.7 (Anthropic) [31]. Each strategy added prompt components while preserving the core evaluation structure, as summarized in Table 3. Initial prompt engineering was carried out with GPT-4o as the development model. We employed iterative refinements in which preliminary CoT prompts were compared against expert scores, and observed discrepancies were used to improve alignment. Once coherence with expert evaluation patterns was achieved, the same three strategies were applied to other models to assess cross-model robustness.

During cross-model adaptation, we maintained an identical logical structure and evaluation approach across all tested models, adjusting only for model-specific formatting requirements. This approach ensured that performance differences between models reflected their inherent capabilities rather than prompt design variations. The detailed implementation of each module was based on the following prompt components:

(1) Module 1 (expert persona) established a medical professor persona with over 20 years of clinical practice and research experience, recognized for precision in evaluating medical documentation. This module emphasized expertise in emergency medicine and diagnostic accuracy, instructing the persona to adopt a rigorous evaluative stance and to carefully review the document and patient information using the five-criteria framework.

(2) Module 2 (error analysis) introduced systematic quantitative error identification before qualitative review. The evaluation covered six error categories: invalid generation error (unwarranted content), non-generation error (missing necessary information), information error (inaccurate details such as values or terminology), prompt echoing error (direct repetition of prompt text), content misplacement error (information in the wrong section), and typo error (grammatical mistakes). Structural malformation errors were prioritized, as our prior research showed they most strongly affect clinical evaluations.

(3) Module 3 (clinical insight integration) incorporated expert-derived insights from our previous findings. It emphasized adherence to structured clinical documentation and prioritizes errors in clinical categories. In particular, errors in differential diagnosis, diagnostic planning, and treatment planning were weighted more heavily than errors in other categories, reflecting their strongest correlation with clinical validity. This module aligned evaluations more closely with expert assessment patterns by emphasizing clinically critical content.

Table 3. Overview of the three-module chain-of-thought framework (MEDIVAL)

Module	Key component
Expert persona	Medical professor persona with extensive experience
	Strict evaluation tendency
	Five criteria assessment framework
Error analysis	Six error categories: invalid generation, non-generation, information, prompt echoing, content misplacement error, typo
	Quantitative evaluation before qualitative assessment
Clinical insight	Integration of expert evaluation patterns
	Emphasis on document structure
	Priority on differential diagnosis, diagnosis, and treatment plan

MEDIVAL, Medical Documentation Validation.

Three progressive CoT strategies

Based on the prompt modules described above, we implemented three progressively sophisticated evaluation strategies, as shown in Table 4. All strategies concluded with a standardized instruction requesting JSON-formatted output. Each model produced scores across the five criteria (appropriateness, accuracy, structure/format, conciseness, and clinical validity) using a 5-point Likert scale, accompanied by detailed comments. The prompt consistently ended with "Let's think step by step" to encourage systematic reasoning.

Persona-based strategy

This baseline strategy applied only module 1, focusing on qualitative, expert-like evaluation. The LLM adopted a medical expert persona to assess documents across the five criteria, producing a structured but primarily subjective assessment comparable to traditional expert review.

Error-enhanced strategy

This enhanced strategy combined modules 1 and 2, integrating systematic error identification prior to qualitative review. Our

Table 4. Relationship between modules and strategies in the chain-of-thought framework

Strategy	Incorporated module
Persona-based strategy	Module 1
Error-enhanced strategy	Modules 1 and 2
Insight-integrated strategy	Modules 1, 2, and 3

Module 1, expert persona. Module 2, error analysis. Module 3, clinical insight.

previous findings showed that structural malformation errors had the strongest negative effect on clinical evaluations ($r = 0.654$, $P < 0.001$), making this quantitative pre-analysis especially important for robust assessment.

Insight-integrated strategy

The most advanced strategy incorporated all three modules, adding expert-derived insights from prior research. This approach emphasized differential diagnosis, diagnostic planning, and treatment planning—elements shown to have the strongest correlation with overall clinical validity ($r = 0.698$, $P < 0.001$). By embedding these clinical priorities, the LLM's evaluation more closely aligned with expert assessment patterns.

Cross-model implementation

We implemented these three CoT strategies across four state-of-the-art LLM architectures:

(1) GPT-4o [28]: OpenAI's multimodal model, featuring integrated vision-language capabilities and optimized latency.
(2) GPT-4.1 [29]: An advanced text-only variant in the GPT-4 series, optimized for long-context reasoning and instruction-following performance.
(3) Claude-3.5 [30]: Anthropic's conversational AI developed with Constitutional AI principles focusing on harmlessness and helpfulness.
(4) Claude-3.7 [31]: Anthropic's advanced model with significantly enhanced reasoning capabilities for complex evaluation tasks.
(5) Other contemporary LLMs were not included in this study due to constraints related to accessibility, computational resources, and cost.

All models used identical parameter settings (temperature = 0, top_p = 1.0), with prompts adapted only for formatting requirements while preserving logical structure. This approach ensured that any observed performance differences reflected model capabilities rather than prompt variation.

Evaluation process implementation

The evaluation process followed a systematic protocol:

(1) Document preparation: All 33 ED records were standardized in format and anonymized to prevent bias.
(2) Prompt construction: For each strategy, we constructed a single general prompt for comparison across models.
(3) Evaluation execution: Each model evaluated all 33 documents under all three strategies, yielding 396 evaluations ($33 \times 4 \times 3$). Each evaluation was repeated four times to assess reproducibility, resulting in 1,584 assessment runs. The repetition mirrored human evaluation conditions, in which four clinicians independently reviewed each record, enabling direct comparison of interrater reliability (ICC) between human and model outputs.
(4) Output standardization: All model outputs were processed to extract standardized JSON data containing the five criteria scores and supporting rationales. This standardization enabled direct comparison across models and strategies.
(5) Quality control: We validated JSON outputs to ensure structural consistency and completeness across all evaluations. Invalid or malformed outputs (less than 0.5% of total) were rerun until valid results were obtained.

All evaluations were performed via application programming interface (API) calls: OpenAI API for GPT-4o and GPT-4.1, and Anthropic Claude API for Claude-3.5 and Claude-3.7. Each model-strategy combination generated standardized outputs with five 5-point Likert scores (maximum total, 25) and supporting justifications.

Statistical analysis

To evaluate the performance of our framework, we conducted a comprehensive statistical analysis using Python ver. 3.11 (Python Software Foundation), with pandas ver. 2.2.0, numpy ver. 1.26.3, scipy ver. 1.12.0, statsmodels ver. 0.14.1, and pingouin ver. 0.5.5. Visualizations were created with matplotlib ver. 3.8.2 and seaborn ver. 0.13.1. All statistical tests were two-sided, with $P < 0.05$ after correction considered significant.

Correlation analysis

We compared mean scores between expert and LLM evaluations using Spearman rank correlation coefficient, selected for its nonparametric properties and robustness to outliers. This makes it particularly appropriate for ordinal data such as Likert scale scores [32,33]. Separate correlation coefficients were calculated for each of the five criteria (appropriateness, accuracy, structure/format, conciseness, and clinical validity) to determine which aspects of medical documentation improved most under progressive strategies. Analyses were conducted independently for each model and prompt strategy to capture nuanced differences in human-model alignment.

Comparative analysis

To examine differences between human and model evaluations, we analyzed score discrepancies across strategies. The Friedman test [34,35] was used as a nonparametric alternative to repeat-

ed-measures analysis of variance, appropriate when normality assumptions are not met. Because each prompt strategy was applied to the same set of documents, the Friedman test also accounted for the paired data structure.

Pairwise comparisons were then performed with the Wilcoxon signed rank test [36], with the Bonferroni correction [37] to adjust for multiple comparisons. This approach was chosen to robustly control the family-wise error rate while respecting the repeated-measures design. Effect sizes were expressed as rank-biserial correlations, which range from −1 to 1 and provide an interpretable measure of both strength and direction of paired differences. Given the modest sample size and potential deviations from normality, nonparametric methods were consistently applied. Boxplots and bar charts were generated to visualize trends across models and strategies.

Reproducibility analysis

Reproducibility was evaluated using the ICC, mean absolute error (MAE), and standard deviation (SD) across four separate evaluation rounds for each document–model combination. We employed ICC(3,k)—a two-way mixed effects model (consistency, average measures)—because it is most suitable when the same evaluator (LLM) performs repeated assessments on fixed items [38]. This analysis was performed separately for each LLM to compare evaluation stability across models.

RESULTS

Correlation analysis

Our analysis showed consistent improvements in correlation with expert evaluations across LLMs as strategies progressed from persona-based to insight-integrated approaches (Fig. 2). Claude-3.7 achieved the strongest overall correlation using the insight-integrated strategy ($r = 0.712$, $P < 0.001$), followed closely by GPT-4o ($r = 0.702$, $P < 0.001$). GPT-4.1 exhibited the largest relative improvement, with a 43.3% increase in correlation from the persona-based ($r = 0.457$, $P = 0.007$) to the insight-integrated approach ($r = 0.655$, $P < 0.001$).

This consistent improvement across architecturally different models suggests that our progressive CoT strategies provide generalizable benefits regardless of the underlying model architecture. However, models with stronger baselines (Claude-3.7 and Claude-3.5) showed smaller relative gains, whereas models with lower starting correlations (GPT-4.1) demonstrated greater improvements.

Comparative analysis

Our analysis of absolute score differences between LLMs and human experts revealed consistent alignment trends across model-strategy combinations (Fig. 3). All LLMs tended to assign higher scores than human experts, with mean differences ranging from 5.37 to 9.00 points on a 25-point scale. Importantly, the insight-integrated strategy consistently reduced these gaps. GPT-4.1 under the insight-integrated strategy showed the smallest mean difference (5.37 points), achieving the closest alignment with human evaluations despite a moderate correlation coefficient.

Boxplot analysis indicated that the insight-integrated strategy reduced both median score differences and variance in many cases, suggesting more consistent alignment with human expert judgment.

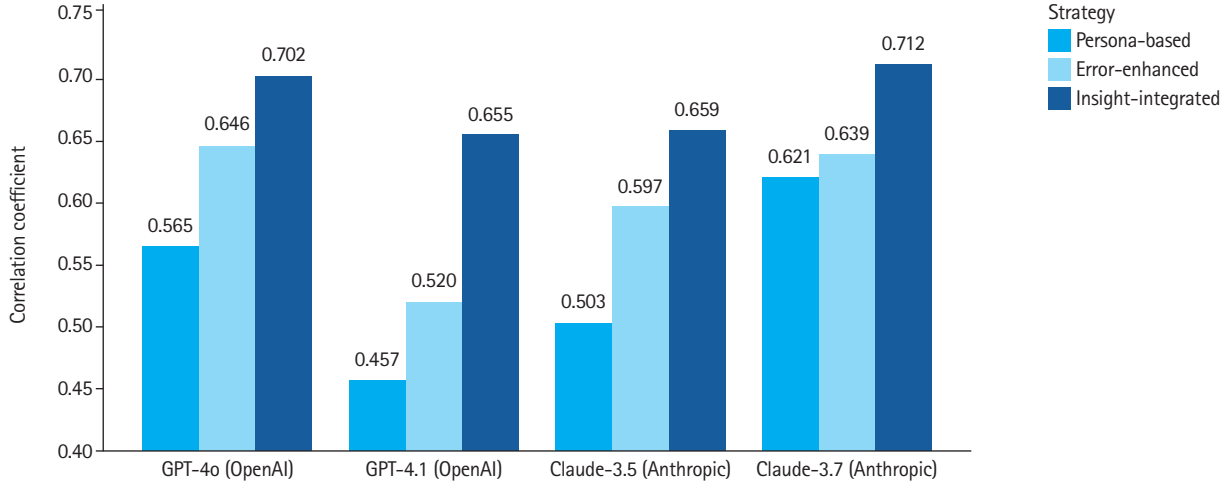


Fig. 2. Agreement between models and clinical experts. All models demonstrated progressive improvement from persona-based to insight-integrated approaches. Claude-3.7 achieved the strongest correlation with expert ratings (r=0.712), followed closely by GPT-4o (r=0.702). All correlations were statistically significant (P<0.001 for most comparisons; GPT-4.1 persona-based P=0.007).

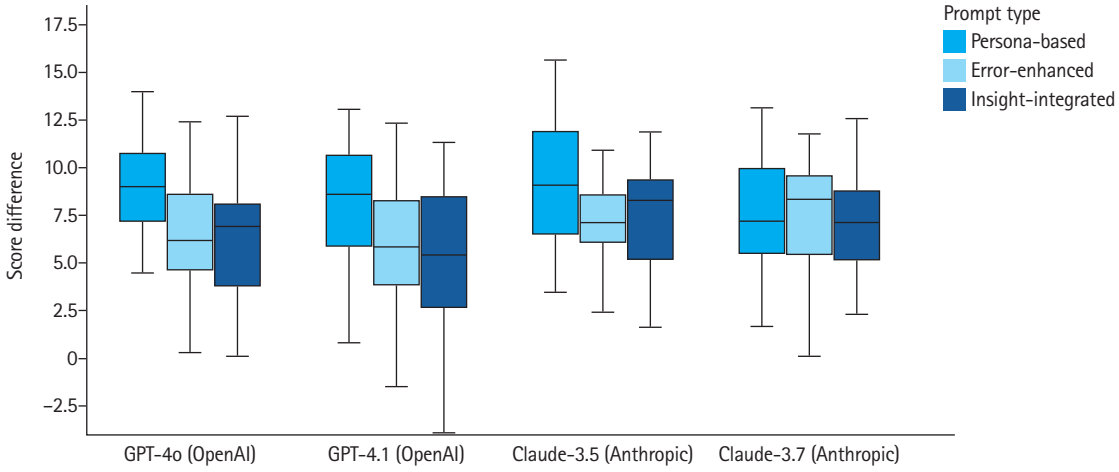


Fig. 3. Score differences between large language model and human expert evaluations on a 25-point scale. Lower values represent closer alignment with expert assessments. The effect of prompt strategies on score differences varied across models. GPT-4.1 with the insight-integrated strategy achieved the closest overall alignment with expert assessments (mean difference, 5.37 points), while error-enhanced strategies yielded smaller differences in GPT-4o and Claude-3.5.

The Friedman test confirmed significant differences in model-human score discrepancies across strategies ($\chi^2(2) = 48.39$, $P < 0.001$). Pairwise Wilcoxon signed rank tests with Bonferroni correction showed that both error-enhanced and insight-integrated strategies significantly reduced score differences compared to the persona-based strategy (both $P < 0.001$).

Although the error-enhanced and insight-integrated strategies also differed significantly ($P = 0.002$), the effect size was modest, indicating only incremental benefit of insight integration beyond error analysis. Descriptive statistics and pairwise comparison results are presented in Table 5.

Table 5. Pairwise comparison of score differences across evaluation strategies

Strategy comparison	Difference (mean ± SD)[a]	Effect size (r)[b]	P-value[c]
Error-enhanced vs. Insight-integrated	0.084 ± 0.253	0.412	0.002
Error-enhanced vs. Persona-based	−0.278 ± 0.305	−1.000	< 0.001
Insight-integrated vs. Persona-based	−0.363 ± 0.417	−0.987	< 0.001

[a]Mean difference indicates the average of per-document score differences. [b]Rank-biserial correlation (range, −1 to 1). [c]Wilcoxon signed rank test with Bonferroni correction.

Detailed analysis of evaluation criteria

Because GPT-4.1 showed the greatest relative improvement across strategies (Fig. 4), we examined its performance across evaluation criteria to illustrate the impact of progressive CoT prompting.

Appropriateness showed the most remarkable improvement, increasing from $r = 0.289$ (persona-based) to $r = 0.610$ (insight-integrated), a 111% increase. This suggests that structured reasoning markedly strengthens GPT-4.1's ability to maintain a professional tone and adhere to documentation standards.

Clinical validity also demonstrated substantial improvement from $r = 0.329$ to $r = 0.603$ (83.3% increase), indicating that the insight-integrated strategy significantly enhances the model's ability to evaluate sound clinical reasoning and diagnostic accuracy in medical documentation.

Structure/format showed the second-largest relative improvement, increasing from $r = 0.439$ to $r = 0.584$ (33% increase), suggesting that GPT-4.1 particularly benefits from guidance in evaluating document organization and logical flow.

Accuracy began with a relatively strong baseline ($r = 0.552$) and improved modestly with insight integration ($r = 0.608$, 10.1% increase). However, the error-enhanced strategy yielded greater improvement ($r = 0.612$, 10.9% increase), implying that systematic error detection is particularly effective for factual correctness in medical documentation.

The clear stepwise improvement across nearly all criteria demonstrates that each added module contributes meaningfully to evaluation quality. GPT-4.1 displayed the most distinct progression among all tested models, suggesting it is especially responsive to structured reasoning prompts.

Reproducibility analysis

To test reliability, we calculated ICC values across four evaluation rounds for each model-strategy combination (Table 6). All combinations showed exceptionally high reproducibility, with ICC values ranging from 0.919 to 0.998. Claude-3.5 demonstrated the highest reproducibility (ICC, 0.997–0.998), while GPT-4o achieved

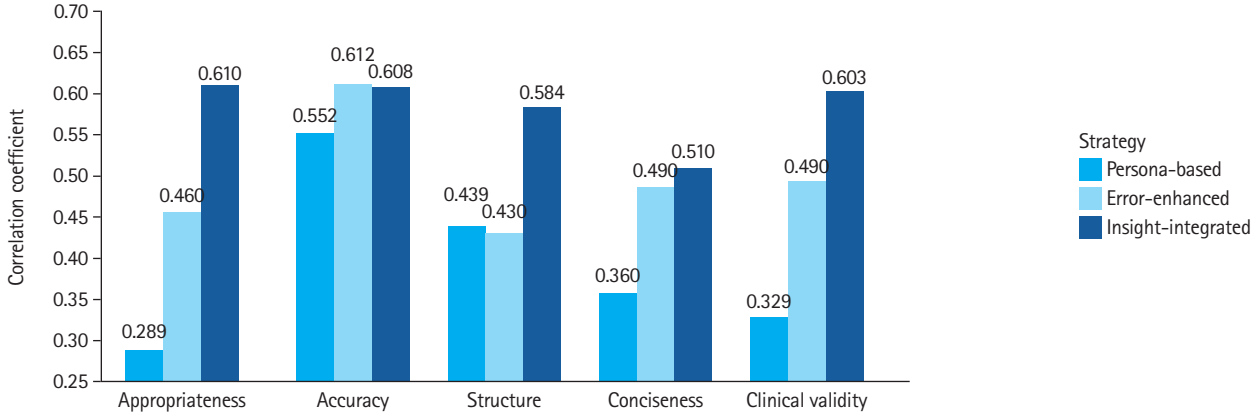


Fig. 4. GPT-4.1 (OpenAI) performance across evaluation criteria and strategies. The insight-integrated approach produced the most dramatic improvement in appropriateness (r=0.610) and substantial gains in clinical validity (r=0.603). All reported correlations were statistically significant (P<0.001).

Table 6. Reproducibility analysis across models and strategies (33 documents, 4 repeated evaluations)

Model	ICC(3,k)	F-test (df1,df2)	P-value	95% CI	MAE	SD
GPT-4o (OpenAI)						
Persona-based	0.986	72.3 (32, 96)	<0.001	0.98–0.99	0.20	0.26
Error-enhanced	0.993	134.4 (32, 96)	<0.001	0.99–1.00	0.26	0.33
Insight-integrated	0.990	95.6 (32, 96)	<0.001	0.98–0.99	0.36	0.46
GPT-4.1 (OpenAI)						
Persona-based	0.990	102.7 (32, 96)	<0.001	0.98–0.99	0.22	0.29
Error-enhanced	0.991	113.3 (32, 96)	<0.001	0.98–1.00	0.44	0.55
Insight-integrated	0.990	103.3 (32, 96)	<0.001	0.98–0.99	0.49	0.64
Claude-3.5 (Anthropic)						
Persona-based	0.997	343.6 (32, 96)	<0.001	1.00–1.00	0.04	0.06
Error-enhanced	0.998	512.8 (32, 96)	<0.001	1.00–1.00	0.09	0.11
Insight-integrated	0.997	388.1 (32, 96)	<0.001	1.00–1.00	0.12	0.17
Claude-3.7 (Anthropic)						
Persona-based	0.919	12.4 (32, 96)	<0.001	0.86–0.96	1.25	1.66
Error-enhanced	0.949	19.6 (32, 96)	<0.001	0.91–0.97	1.42	1.89
Insight-integrated	0.969	32.3 (32, 96)	<0.001	0.95–0.98	1.32	1.76

ICC, intraclass correlation coefficient; CI, confidence interval; MAE, mean absolute error; SD, standard deviation.

peak reproducibility under the insight-integrated strategy (ICC, 0.990).

MAE varied by model, with Claude-3.5 showing the lowest errors (range, 0.04–0.12) and Claude-3.7 the highest (range, 1.25–1.42). Interestingly, the insight-integrated strategy occasionally produced slightly higher MAE values, suggesting its more complex reasoning could introduce minor variability. Nevertheless, the extremely high ICC values indicate this variability remained minimal.

For comparison, human expert evaluations showed good reliability, with ICC values ranging from 0.653 to 0.887 across criteria, and the highest consistency for clinical validity (ICC, 0.887). While human reliability was strong, all LLM model-strategy combinations in this study demonstrated higher reproducibility, underscoring the potential advantage of LLM-based evaluation frameworks in ensuring consistent assessments.

DISCUSSION

Enhancing LLM evaluation through progressive CoT strategies

Our three-module CoT evaluation framework demonstrated the value of progressively integrating reasoning strategies into LLM-based assessment systems for medical documentation. The improvement patterns observed across different architectures provide evidence that structured reasoning can enhance alignment with expert evaluations regardless of baseline capability. While correlation coefficients generally improved from persona-based to insight-integrated strategies, the magnitude and pattern of improvement varied across models. GPT-4.1 was particularly no-

table, showing the most distinct stepwise improvement (43.3% overall increase). This finding indicates that models with moderate baseline performance may gain the most from structured CoT techniques. In this way, our framework can potentially "level the playing field" across LLMs by enabling even smaller or less specialized models to approach the evaluation quality of larger models when guided by structured strategies.

Performance also varied across evaluation criteria. Accuracy benefited substantially from structured reasoning, suggesting that stepwise analysis enhances factual correctness assessment. Likewise, clinical validity improved significantly under the insight-integrated strategy, indicating that this approach effectively captures the complex reasoning patterns used by human experts in documentation review.

These improvements carry direct clinical implications. Closer alignment with expert judgment strengthens the reliability of LLMs as support tools for documentation review, enabling earlier error detection while safeguarding patient safety. Importantly, gains in clinical validity imply that progressive CoT strategies can help LLMs approximate expert-level reasoning, yielding not only statistical improvements but also clinically meaningful improvements.

Interpretive challenges in human–model evaluation alignment

Our analysis showed that the insight-integrated strategy consistently narrowed the gap between LLM and human evaluations across all models. However, evaluation performance peaked with the error-enhanced strategy rather than the insight-integrated strategy. The Wilcoxon signed rank analysis confirmed that the difference between these two strategies, though statistically significant, was modest in magnitude.

This suggests that objectively measurable features, such as information density and structural completeness, are best captured through error-enhanced strategies, while more subjective elements, such as clinical reasoning quality, may be better reflected by insight integration. Future frameworks may need to balance objective and subjective evaluation components depending on their intended use.

Notably, all LLMs scored medical documents higher than human experts, with mean differences ranging from 5.37 to 9.00 points on a 25-point scale. This finding aligns with our earlier study, which showed that human experts tend to focus more critically on categories such as diagnoses and treatment plans, as well as structural errors. These areas often drove harsher scoring from experts. In contrast, LLMs are typically tuned to align with general human preferences [39], potentially resulting in outputs that are more superficially fluent but less rigorously scrutinized for critical clinical content. This tendency may explain why LLM-generated outputs appear relatively more favorable under surface-level evaluation but might lack the depth required for thorough clinical documentation assessment.

Future applications and system integration

Our findings indicate that frameworks like MEDIVAL could play an important role in clinical documentation workflows. One immediate application is as a prescreening system that flags records requiring expert review, reducing routine workload. In hybrid workflows, models could also direct clinician attention to high-risk sections, such as differential diagnoses or treatment plans, where errors are most consequential. Additionally, criterion-specific feedback may serve an educational role in clinical training.

LLMs exhibited even higher interrater reliability than was observed for human experts (ICC, 0.653–0.887 across evaluation criteria), highlighting the potential advantages of adopting LLMs as evaluators. With the exception of Claude-3.7, most models achieved MAE and SD values below 1 point, suggesting that progressive CoT strategies not only improve relative evaluation but also calibrate absolute scoring more closely to human judgment. Such calibration is essential in clinical settings, where understanding the degree of deviation from acceptable standards is critical for patient safety [40].

Although this study focused on text-based documentation, our framework could extend to emerging modalities such as voice-based electronic medical records and broader ambient AI systems. With the increasing adoption of voice recognition in healthcare [41], future research should explore how structured evaluation frameworks can adapt to challenges in these modalities, including recognition errors, contextual ambiguity, and automated coding accuracy.

Ethical and operational considerations

Despite the demonstrated benefits and promising potential, careful consideration is required when deploying LLM-based evaluators in practice [42,43]. The consistent tendency of LLMs to score documents more favorably than human experts underscores the need for calibration and oversight. While our progressive CoT strategies narrow this scoring gap, they do not eliminate it. Human oversight therefore remains essential, particularly in high-stakes clinical contexts where documentation directly informs patient care.

Transparency regarding model limitations, data provenance,

and evaluation methodologies is critical for building user trust [44,45]. Although this study used only synthetic data without patient identifiers, future applications involving real-world documentation must adhere strictly to institutional governance policies and privacy regulations. Moreover, healthcare systems must evaluate fairness and robustness across diverse clinical contexts to prevent propagation of biases in LLM assessments.

Beyond these ethical and governance issues, practical implementation requires addressing computational efficiency, interoperability with existing EHR systems, and clinician acceptance. These operational factors will be decisive in determining the feasibility of integrating LLM-based evaluators into routine workflows and should be prioritized in translational research efforts.

Collectively, our findings support the gradual integration of LLM-based evaluators into clinical workflows, beginning with expert-augmented decision support and progressively advancing toward greater automation as reliability and trust are established.

Limitations

Several limitations must be considered when interpreting our results. First, prompts were optimized specifically for GPT-4o, reflecting relative rather than absolute model capabilities. Tailored optimization for other models could further improve their performance beyond what was observed here.

Second, our analysis was based on the complete dataset of 33 documents, precluding a formal sample size calculation. Instead, each document underwent two independent reviews by four senior clinicians across five criteria, generating 1,320 structured expert judgments. This dense annotation enabled robust reliability analyses (ICC, test-retest) that would not have been feasible with a larger but less thoroughly annotated dataset. Nonetheless, broader validation will require large-scale testing across diverse clinical settings and documentation types.

Finally, medical documentation standards vary substantially across clinical departments and stages of patient care. The present study did not define a universally applicable "gold standard" medical record, nor did it address documentation requirements across different care contexts and timepoints. Future work should therefore develop adaptable methodologies and evaluation tools that accommodate department-specific standards and temporal variation.

Conclusions

This study demonstrates that progressive CoT strategies substantially enhance the ability of LLMs to evaluate medical documentation in alignment with human expert judgment, thereby reducing the burden of manual evaluation. Future research should validate this framework in varied documentation contexts and pursue practical implementation steps, including pilot deployments, calibration protocols, and clinician training, to ensure safe and effective integration of LLM-based evaluation into healthcare workflows.

ARTICLE INFORMATION

Author contributions

Conceptualization: DC, TK; Data curation: JS, WCC; Formal analysis: DC, MK; Investigation: SH, HC; Methodology: DC, TK; Project administration: TK; Software: DC; Supervision: TK; Writing–original draft: DC, JS; Writing–review & editing: all authors. All authors read and approved the final manuscript.

Conflicts of interest

Won Chul Cha and Taerim Kim are editorial board members of this journal, but were not involved in the peer reviewer selection, evaluation, or decision process of this article. The authors have no other conflicts of interest to declare.

Funding

The authors received no financial support for this study.

Acknowledgments

The authors would like to thank all participants who contributed to the original data collection for this study.

Data availability

Data analyzed in this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Kuhn T, Basch P, Barr M, Yackel T. Clinical documentation in the 21st century: executive summary of a policy position paper from the American College of Physicians. Ann Intern Med 2015;162:301–3.
2. Sanderson AL, Burns JP. Clinical documentation for intensivists: the impact of diagnosis documentation. Crit Care Med 2020;48:579–87.
3. Lorenzetti DL, Quan H, Lucyk K, et al. Strategies for improving physician documentation in the emergency department: a systematic review. BMC Emerg Med 2018;18:36.
4. Shala DR, Jones A, Fairbrother G, Thuy Tran D. Completion of

electronic nursing documentation of inpatient admission assessment: insights from Australian metropolitan hospitals. Int J Med Inform 2021;156:104603.
5. Ahn M, Choi M, Kim Y. Factors associated with the timeliness of electronic nursing documentation. Healthc Inform Res 2016;22:270–6.
6. Gesner E, Gazarian P, Dykes P. The burden and burnout in documenting patient care: an integrative literature review. Stud Health Technol Inform 2019;264:1194–8.
7. Moy AJ, Schwartz JM, Chen R, et al. Measurement of clinical documentation burden among physicians and nurses using electronic health records: a scoping review. J Am Med Inform Assoc 2021;28:998–1008.
8. Gesner E, Dykes PC, Zhang L, Gazarian P. Documentation burden in nursing and its role in clinician burnout syndrome. Appl Clin Inform 2022;13:983–90.
9. Murad MH, Vaa Stelling BE, West CP, et al. Measuring documentation burden in healthcare. J Gen Intern Med 2024; 39:2837–48.
10. Levy DR, Withall JB, Mishuris RG, et al. Defining documentation burden (DocBurden) and excessive DocBurden for all health professionals: a scoping review. Appl Clin Inform 2024; 15:898–913.
11. Preiksaitis C, Wright KN, Alvarez A, et al. Measuring burnout and professional fulfillment among emergency medicine residency program leaders in the United States: a cross-sectional survey study. Clin Exp Emerg Med 2025;12:76–85.
12. Li LZ, Yang P, Singer SJ, Pfeffer J, Mathur MB, Shanafelt T. Nurse burnout and patient safety, satisfaction, and quality of care: a systematic review and meta-analysis. JAMA Netw Open 2024;7:e2443059.
13. Sloss EA, Abdul S, Aboagyewah MA, et al. Toward alleviating clinician documentation burden: a scoping review of burden reduction efforts. Appl Clin Inform 2024;15:446–55.
14. Holmgren AJ, Hendrix N, Maisel N, et al. Electronic health record usability, satisfaction, and burnout for family physicians. JAMA Netw Open 2024;7:e2426956.
15. Park S, Marquard J, Austin RR, Pieczkiewicz D, Jantraporn R, Delaney CW. A systematic review of nurses' perceptions of electronic health record usability based on the human factor goals of satisfaction, performance, and safety. Comput Inform Nurs 2024;42:168–75.
16. Ding H, Simmich J, Vaezipour A, Andrews N, Russell T. Evaluation framework for conversational agents with artificial intelligence in health interventions: a systematic scoping review. J Am Med Inform Assoc 2024;31:746–61.
17. Wei Q, Yao Z, Cui Y, Wei B, Jin Z, Xu X. Evaluation of ChatGPT-generated medical responses: a systematic review and meta-analysis. J Biomed Inform 2024;151:104620.
18. Satheakeerthy S, Jesudason D, Pietris J, Bacchi S, Chan WO. LLM-assisted medical documentation: efficacy, errors, and ethical considerations in ophthalmology. Eye (Lond) 2025;39: 1440–2.
19. Lee YC. Rethinking artificial intelligence in medicine: from tools to agents. Clin Exp Emerg Med 2025;12:101–3.
20. Goodman KE, Yi PH, Morgan DJ. AI-generated clinical summaries require more than accuracy. JAMA 2024;331:637–8.
21. Seo J, Choi D, Kim T, et al. Evaluation framework of large language models in medical documentation: development and usability study. J Med Internet Res 2024;26:e58329.
22. Hartman V, Zhang X, Poddar R, et al. Developing and evaluating large language model-generated emergency medicine handoff notes. JAMA Netw Open 2024;7:e2448723.
23. Gu J, Jiang X, Shi Z, et al. A survey on LLM-as-a-judge [Preprint]. arXiv:2411.15594; [posted 2024 Nov 23]. https://doi.org/10.48550/arXiv.2411.15594
24. Li D, Jiang B, Huang L, et al. From generation to judgment: opportunities and challenges of LLM-as-a-judge [Preprint]. arXiv:2411.16594; [posted 2024 Nov 25]. https://doi.org/10.48550/arXiv.2411.16594
25. Seo J, Choi D, Cha W, Kim T. LLM-based medical document evaluation: integrating human expert insights. Stud Health Technol Inform 2025;329:1029–33.
26. Naver Cloud. HyperCLOVA X [Internet]. Naver Cloud; [cited 2025 May 1]. Available from: https://www.ncloud.com/solution/featured/hyperclovax
27. Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models [Preprint]. arXiv:2201.11903; [posted 2022 Jan 28]. https://doi.org/10.48550/arXiv.2201.11903
28. OpenAI, Hurst A, Lerer A, et al. GPT-4o system card [Preprint]. arXiv:2410.21276; [posted 2024 Oct 25]. https://doi.org/10.48550/arXiv.2410.21276
29. OpenAI. Introducing GPT-4.1 in the API [Internet]. OpenAI; 2025 [cited 2025 May 1]. Available from: https://openai.com/index/gpt-4-1/
30. Anthropic [Internet]. Anthropic; [cited 2025 May 1]. Available from: https://www.anthropic.com
31. Anthropic. Claude 3.7 Sonnet system card. Anthropic; 2025.
32. Hauke J, Kossowski T. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaest Geogr 2011;30:87–93.

33. Siegel S. Nonparametric statistics. Am Stat 1957;11:13–9.
34. Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J Am Stat Assoc 1937;32:675–701.
35. Friedman M. A comparison of alternative tests of significance for the problem of m rankings. Ann Math Statist 1940;11:86–92.
36. Wilcoxon F. Individual comparisons by ranking methods. In: Kotz S, Johnson NL, editors. Breakthroughs in statistics. Springer; 1992. p. 196–202.
37. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995;310:170.
38. Ionan AC, Polley MY, McShane LM, Dobbin KK. Comparison of confidence interval methods for an intra-class correlation coefficient (ICC). BMC Med Res Methodol 2014;14:121.
39. Liu W, Wang X, Wu M, et al. Aligning large language models with human preferences through representation engineering [Preprint]. arXiv:2312.15997; [posted 2023 Dec 26]. https://doi.org/10.48550/arXiv.2312.15997
40. Fragiadakis G, Diou C, Kousiouris G, Nikolaidou M. Evaluating human-AI collaboration: a review and methodological framework [Preprint]. arXiv:2407.19098; [posted 2024 Jul 9]. https://doi.org/10.48550/arXiv.2407.19098
41. Owens LM, Wilda JJ, Hahn PY, Koehler T, Fletcher JJ. The association between use of ambient voice technology documentation during primary care patient encounters, documentation burden, and provider burnout. Fam Pract 2024;41:86–91.
42. Li F, Ruijs N, Lu Y. Ethics & AI: a systematic review on ethical concerns and related strategies for designing with AI in healthcare. AI 2023;4:28–53.
43. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022;28:31–8.
44. Uygun Ilikhan S, Özer M, Tanberkan H, Bozkurt V. How to mitigate the risks of deployment of artificial intelligence in medicine? Turk J Med Sci 2024;54:483–92.
45. Okada Y, Ning Y, Ong ME. Explainable artificial intelligence in emergency medicine: an overview. Clin Exp Emerg Med 2023;10:354–62.

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):65-73
https://doi.org/10.15441/ceem.24.344

Original Article

Skull fractures may be associated with hyperfibrinolysis in patients with isolated traumatic brain injury

Gaku Fujiwara[1], Naoto Shiomi[2]

[1]Department of Neurosurgery, Saiseikai Shiga Hospital, Imperial Gift Foundation Inc, Ritto, Japan
[2]Department of Critical and Intensive Care Medicine, Shiga University of Medical Science, Shiga, Japan

pISSN: 3022-1811
eISSN: 2383-4625

Objective To identify relationships between skull fracture (SF) and hyperfibrinolysis (HF) among patients with isolated traumatic brain injury (TBI).

Methods This was a retrospective cohort study based on a nationwide neurotrauma database in Japan. Adult patients with isolated TBI (head Abbreviated Injury Scale [AIS] >2, any other AIS <3) and who were registered in the multicenter neurotrauma registry from 2015 to 2017 were included. To examine the relationship between SF and HF, we conducted multivariable logistic regression analyses to calculate the adjusted odds ratios (aORs) with their 95% confidence intervals (CIs) for HF. HF was defined as a D-dimer level ≥38 mg/L on arrival based on a previous study.

Results A total of 335 patients were enrolled and the median age of the cohort was 64 years (interquartile range, 44–76 years). HF was observed in 161 patients (48.1%). The association of SF with HF yielded an aOR of 4.78 (95% CI, 2.71–8.42) compared to non-SF in multivariable logistic regression analysis. In addition, the associations of skull base fracture, skull vault fracture, and combination of skull base and vault fracture with HF yielded the corresponding aORs of 3.60 (95% CI, 1.20–10.81), 4.99 (95% CI, 2.63–9.44), and 4.84 (95% CI, 2.41–9.72), respectively, relative to non-SF.

Conclusion This multicenter observational study demonstrated the association of SF with HF in patients with isolated TBI.

Keywords Traumatic brain injury; Skull fracture; Hyperfibrinolysis; Coagulopathy; D-dimer

Capsule Summary

What is already known

Hyperfibrinolysis (HF), indicated by elevated D-dimer levels, is a well-established complication in trauma-induced coagulopathy and is associated with traumatic brain injury (TBI). Skull fractures (SFs) are often observed in TBI and may indicate severe brain tissue injury. Prior studies have suggested a relationship between SF and coagulopathy, but no studies have explicitly reported an association between SF and HF.

What is new in the current study

This multicenter cohort study demonstrates a significant independent association between SF and HF in patients with isolated TBI. Specifically, the presence of skull vault fractures, skull base fractures, or a combination of both, was found to be strongly associated with elevated D-dimer levels on admission. These findings suggest that SF may serve as an early predictor of HF, providing valuable information for trauma management strategies.

Received: October 23, 2024
Revised: January 26, 2025
Accepted: March 18, 2025

Correspondence to:
Gaku Fujiwara
(gfujiwara-tuk@umin.ac.jp)

How to cite this article:
Fujiwara G, Shiomi N. Skull fractures may be associated with hyperfibrinolysis in patients with isolated traumatic brain injury. Clin Exp Emerg Med 2026;13(1):65-73. https://doi.org/10.15441/ceem.24.344

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

© 2026 The Korean Society of Emergency Medicine

INTRODUCTION

Trauma patients often suffer from coagulopathy, which can result in mortality resulting from the trauma [1–4]. Hyperfibrinolysis (HF), characterized by elevated D-dimer levels, is a well-recognized manifestation of trauma-induced coagulopathy [3,5–9]. In the case of traumatic brain injury (TBI), D-dimer levels not only reflect HF but also serve as an important prognostic indicator [10,11]. HF is a critical factor in TBI as it is associated with the progression of intracranial hemorrhage and poorer prognosis [12]. However, the extent to which different types of morphology in TBI are correlated with HF remains unclear.

Skull fractures (SFs) are commonly observed in TBI and may be an objective indicator of the severity of head impact and primary brain injury, as they occur when force is sufficient to cause skull fracture [13,14]. Brain tissue injury-induced release of tissue factor has been implicated in promoting HF [1,5,15–17]. TBI with SF may result in more severe brain tissue damage than TBI alone. Previous research has shown an independent relationship between SF and prognosis [18], and another report indicates that D-dimer levels are associated with SF in pediatric TBI [19]. In a previous study, we found that depressed SF was associated with severe hypofibrinogenemia, which means SF may induce HF and cause fibrinogen depletion in TBI settings [20]. However, no studies have explicitly reported an association between SF and HF.

We hypothesize that the presence of SF is correlated with HF, and early detection and intervention of HF in TBI are crucial. Therefore, it is necessary to investigate this relationship to improve the quality of trauma care. In this study we investigated the relationship between SF and elevated D-dimer levels in patients with TBI on admission and determine the relationships between injury morphology and HF.

METHODS

Ethics statement

Permission to analyze the Japan Neurotrauma Data Bank (JNTDB) database was obtained from the Japan Society of Neurotraumatology before starting this study. The study protocol was approved by the Institutional Ethics Review Board of Yamaguchi University (No. H26-36), which was the core hospital of this registry. The requirement for patient consent was waived because only anonymized data were used. This study was conducted in compliance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting guidelines [21].

Study design and setting

The data were obtained from the JNTDB dataset, which contains information collected from 33 hospitals across Japan from April 2015 to March 2017. The JNTDB is a nationwide, multicenter, prospective, observational trauma registry. The data were compiled by administrators based on in-hospital electronic medical charts. Detailed descriptions of the JNTDB have been published previously [22,23]. The following patients were included in the JNTDB: (1) all age groups; (2) patients with a Glasgow Coma Scale (GCS) score of 8 or less at the time of admission or who deteriorated to a GCS score of 8 or less during the clinical course; and (3) patients who underwent neurosurgical intervention as treatment for trauma (including trepanation but excluding surgery for chronic subdural hematoma). Details of the JNTDB are provided in Suppl. 1.

Study participants

Adult trauma patients (≥16 years) registered in the JNTDB dataset with 1,345 patients were considered eligible for inclusion. We included patients with isolated TBI, defined as head Abbreviated Injury Scale (AIS) >2 and any other AIS <3. We excluded patients younger than 16 years and those who developed cardiac arrest on arrival at the hospital. We also excluded patients who had received antithrombotic therapy before the injury, those who were transported more than 3 hours after injury, and those who were transported via other hospitals. Moreover, we excluded patients with missing information on the D-dimer level.

Data collection, variables, and potential bias

We collected and described the following clinical information from the JNTDB: sex, age, time of transport, vital signs (systolic blood pressure, heart rate, body temperature) and GCS score at hospital arrival, AIS of the head, Injury Severity Score (ISS), and in-hospital mortality. We also collected imaging studies for the presence of SF, which was subdivided into skull base fracture (base-SF), skull vault fracture (vault-SF), and merging both fractures (both-SF), type of intracranial injury and thickness or diameter of hematoma (subdural hematoma, epidural hematoma, cerebral contusion, subarachnoid hemorrhage [SAH], intraventricular hemorrhage [IVH]), Trauma Coma Data Bank (TCDB) classification (diffuse injury I–II, diffuse injury III–IV, evacuated/nonevacuated mass), and ISS. The AIS, ISS, and TCDB classification details are described in Suppl. 2. The TCDB classification is divided into diffuse injury I to IV, evacuated mass, and nonevacuated mass [24,25]. These were further categorized as diffuse injury I and II, diffuse injury III and IV, evacuated mass, and nonevacuated mass

for analysis. In addition, we collected coagulation parameters, including D-dimer, on arrival.

Definitions

We defined HF as a D-dimer level ≥38 mg/L on arrival, based on a previous multicenter study in Japan reported by Hayakawa et al. [7]. Therefore, we defined the presence or absence of HF using two groups: HF(+) (D-dimer ≥38 mg/L) and HF(–) (D-dimer <38mg/L).

Statistical methods

We described the patient characteristics and distribution of patients with and without SF (SF and non-SF groups). We defined the two groups as the explanatory variable and the presence of HF as the objective variable. We performed logistic regression analyses to generate crude odds ratios (ORs) for HF(+) with SF, along with their 95% confidence intervals (CIs), compared to non-SF. For potential confounders, we included sex, age, GCS, type of intracranial injury, and TCDB classification as covariates and performed multivariable logistic regression analyses to calculate the adjusted ORs (aORs) of the groups with and without SF with their 95% CIs. Further, we also performed subgroup analysis to identify any associations with HF. Statistical results were calculated as point estimates with their 95% CIs. Statistical significance was defined as the absence of a 95% CI overlap with the null effect value (OR=1). All statistical analyses were performed using JMP Pro ver. 17 (SAS Institute Inc).

Subgroup analysis

We divided the cohort into a nonelderly group (<65 years) and an elderly group (≥65 years) and evaluated D-dimer levels and HF in each subgroup. This was done to examine how SF might be affected by differences in D-dimer levels according to age as reported in a previous study [26].

Sensitivity analysis

We performed five additional sensitivity analyses to confirm the robustness of our results. Because the definition of HF by D-dimer level is unclear, we performed a sensitivity analysis using different cutoff values. Logistic regression analysis was performed using D-dimer >50 mg/L, which we used as the cutoff for severe HF in our previous study [20]. Considering the complex pathophysiology of coagulopathy in TBI, we conducted an additional analysis in which HF was defined as fibrinogen <150 mg/dL, rather than using D-dimer as the sole criterion. Fibrinogen decreases in TBI due to depletion caused by HF [20,27,28]. Intracranial hematoma was assessed as a binary value of presence or absence in the main analysis, but a continuous variable of hematoma size (thickness and diameter) was used as a confounding variable. To further eliminate the influence of torso trauma, a logistic regression analysis was performed on the cohort that further excluded those with AIS of 2 trauma to other parts of the body other than the head. Logistic analysis was performed by adding ISS, which has been shown to be associated with coagulopathy in TBI, to the variables [29].

RESULTS

Patient characteristics

The study flowchart is shown in Fig. 1. Of 1,345 patients enrolled in the JNTDB, 335 patients were included in the analysis: 202 (60.3%) in the SF group and 133 (39.7%) in the non-SF group. Patient characteristics and clinical data are described in Table 1. More than two-thirds of the patients were male (68.7%), and the median age was 64 years (interquartile range [IQR], 44–76 years). Median GCS score of the patients was 7 (IQR, 4–10). The median D-dimer level was 34.6 mg/L (IQR, 11.2–87.7 mg/L).

HF assessment

The characteristics of patients according to HF(+) or HF(–) are

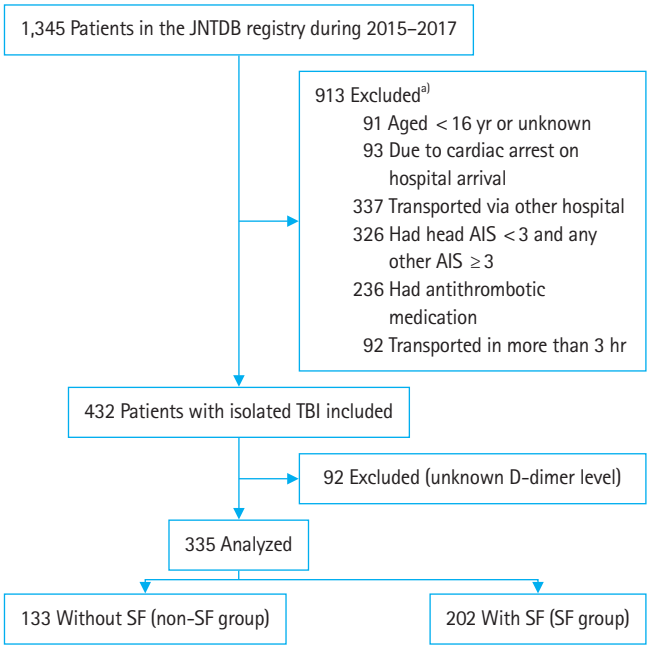


Fig. 1. Flowchart of the patient selection. JNTDB, Japan Neurotrauma Data Bank; AIS, Abbreviated Injury Scale; TBI, traumatic brain injury; SF, skull fracture. [a)]The sum exceeds the total because some individuals met more than one criterion.

described in Suppl. 3. The boxplot of D-dimer levels by each SF is described in Fig. 2. The overall incidence of HF(+) was 48.1% (n = 161). The association of SF with HF(+) yielded aORs of 4.78 (95% CI, 2.71–8.42), relative to non-SF in multivariable logistic regression analysis (Fig. 3). Multivariable logistic regression analyses of the association of base-, vault-, and both-SF with HF(+) yielded the corresponding aORs of 3.60 (95% CI, 1.20–10.81), 4.99 (95% CI, 2.63–9.44), and 4.84 (95% CI, 2.41–9.72), respectively, relative to non-SF (Fig. 3). Crude ORs and aORs of other covariates are described in Suppl. 4. The distribution of D-dimer values and HF evaluated in the nonelderly and elderly group subgroups is shown in Suppl. 5, and the logistic regression analysis results are shown in Table 2.

Table 1. Characteristics of the study participants

Characteristic	Total (n = 335)	SF group (n = 202)	Non-SF group (n = 133)
Sex			
Male	230 (68.7)	143 (70.8)	87 (65.4)
Female	105 (31.3)	59 (29.2)	46 (34.6)
Age (yr)	64 (44–76)	63 (43–73)	66 (43–78)
Transport time (min)	40 (30–52)	40 (30–52)	43 (30–56)
Heart rate (bpm)	86 (74–101)	87 (74–101)	86 (74–99)
Systolic blood pressure (mmHg)	151 (130–178)	149 (127–177)	157 (136–180)
Body temperature (°C)	36.2 (35.8–36.7)	36.1 (35.8–36.5)	36.4 (36–37.0)
GCS score	7 (4–10)	7 (4–11)	6 (4–8)
Type of hematoma			
Subdural hematoma	172 (51.3)	98 (48.5)	74 (55.6)
Epidural hematoma	54 (16.1)	46 (22.8)	8 (6.0)
Contusion	54 (16.1)	40 (19.8)	14 (10.5)
Type of intracranial injury			
Subarachnoid hemorrhage	262 (78.2)	178 (88.1)	84 (63.2)
Intraventricular hemorrhage	52 (15.5)	34 (16.8)	18 (13.5)
Type of SF			
Base fracture	18 (5.4)	18 (8.9)	0 (0)
Vault fracture	105 (31.3)	105 (52.0)	0 (0)
Base and vault fracture	79 (23.6)	79 (39.1)	0 (0)
TCDB classification			
Diffuse injury I–II	78 (23.2)	35 (17.3)	43 (32.3)
Diffuse injury III–IV	31 (9.3)	22 (10.9)	9 (6.8)
Evacuated and non-evacuated mass	226 (67.5)	145 (71.8)	81 (60.9)
Injury Severity Score	25 (16–25)	25 (16–25)	25 (16–25)
Coagulation parameter			
PT-INR	1.05 (0.99–1.14)	1.06 (0.99–1.15)	1.05 (0.98–1.13)
APTT (sec)	28.5 (24.9–33.1)	28.6 (24.7–33.8)	28.4 (25.4–32.3)
Fibrinogen (mg/dL)	233 (178–296)	214 (168–274)	258 (195–320)
D-dimer (mg/L)	34.6 (11.2–87.7)	48.8 (22.9–225.2)	17.9 (5.6–45.6)
Surgical procedure	241 (71.9)	155 (76.7)	86 (64.7)
Outcome			
In-hospital mortality	106 (31.6)	75 (37.1)	31 (23.3)
GOS poor[a] at discharge	219 (65.4)	129 (63.9)	90 (67.7)
HF(+)	161 (48.1)	122 (60.4)	39 (29.3)

Values are presented as number (%) or median (interquartile range).
SF, skull fracture; bpm, beats per minute; GCS, Glasgow Coma Scale; TCDB, Traumatic Coma Data Bank; PT, prothrombin time; INR, international normalized ratio; APTT, activated partial thromboplastin time; GOS, Glasgow Outcome Scale; HF, hyperfibrinolysis.
[a]GOS score 1–3.

Sensitivity analysis

The results of the sensitivity analysis are shown in Table 3. For sensitivity analysis 4, 73 cases with AIS of 2 at any other injury site were excluded, and 262 cases were deemed eligible for analysis. Sensitivity analysis results were similar to those of the main analysis.

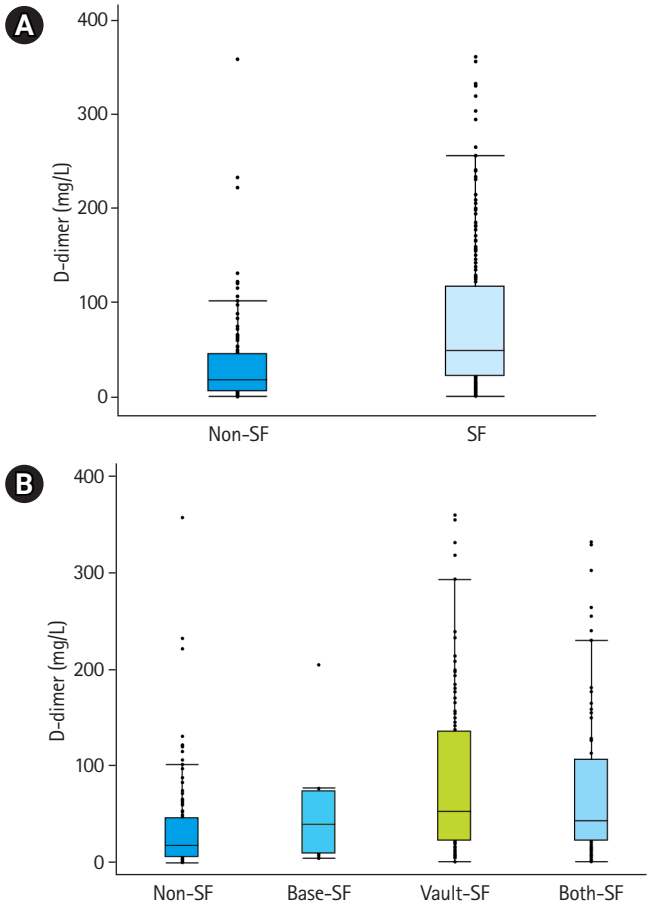


Fig. 2. Boxplot of D-dimer levels in each skull fracture (SF) group. (A) Non-SF and SF groups. (B) Non-SF and subgroups of SF (skull base fracture [base-SF], skull vault fracture [vault-SF], and merging both fractures [both-SF]).

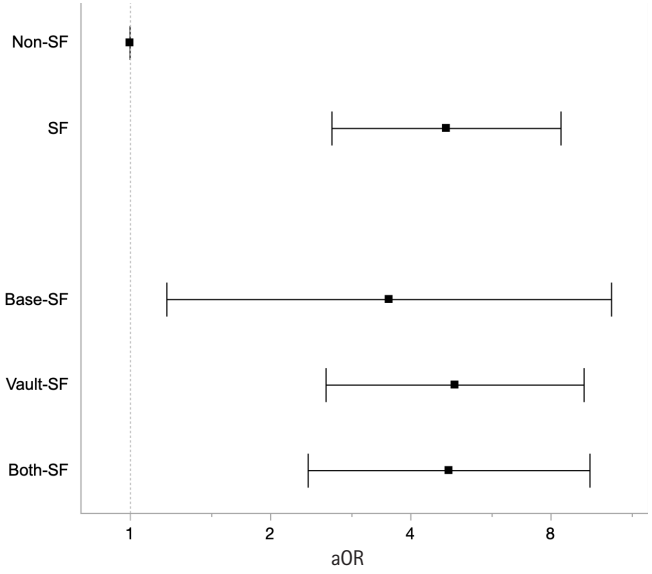


Fig. 3. Forest plot of adjusted odds ratios (aORs) for hyperfibrinolysis (HF) in each skull fracture (SF) group (adjusted for sex, age, Glasgow Coma Scale, type of intracranial injury, the Trauma Coma Data Bank classification). SF in the main analysis includes all types of SF (skull base fracture [base-SF], skull vault fracture [vault-SF], and merging both fractures [both-SF]).

Table 2. Crude ORs and aORs for hyperfibrinolysis of multivariable logistic regression analysis in each subgroup divided by age

Variable	Crude OR (95% CI)	aOR[a] (95% CI)
Nonelderly subgroup (< 65 yr)		
Non-SF	Reference	Reference
SF	3.39 (1.62–7.08)	4.08 (1.71–9.74)
Elderly subgroup (≥ 65 yr)		
Non-SF	Reference	Reference
SF	5.84 (2.96–11.5)	6.05 (2.67–13.7)

OR, odds ratio; aOR, adjusted odds ratio; CI, confidence interval; SF, skull fracture.
[a]Adjusted for sex, age, Glasgow Coma Scale, type of intracranial injury, and the Traumatic Coma Data Bank classification.

DISCUSSION

In this multicenter study including 33 hospitals, patients with SF showed a clear independent association with high D-dimer levels, a finding suggestive of HF. Furthermore, HF was more frequently observed across various SF morphologies, including vault fractures, and a significant association with HF was consistently demonstrated across all SF types compared to the non-SF group.

Our study has several strengths. First, the results of our study were novel: there are few reports that SF is associated with HF, and even fewer reports that it depends on the extent of SF. In a previous study, we reported that SF was associated with in-hospital mortality. It is independently associated with intracranial injury, but we could not attribute this to HF [18]. Another multicenter study reported that the presence of vault-SF was associated with the occurrence of talk and die, and an association is suggested if talk and die is interpreted as being due to increased intracranial hematoma caused by HF. However, no coagulation parameters were reported in the previous study [30]. As described above, although previous reports suggested an association with HF, none explicitly demonstrated it, making the present study results novel.

Table 3. Crude ORs and aORs of multivariable logistic regression analysis in the sensitivity analysis

Variable	Crude OR (95% CI)	aOR[a] (95% CI)
Sensitivity analysis 1		
Non-SF	Reference	Reference
SF	3.51 (2.14–5.77)	4.29 (2.39–7.70)
Sensitivity analysis 2		
Non-SF	Reference	Reference
SF	1.83 (0.94–3.55)	2.37 (1.11–5.10)
Sensitivity analysis 3		
Non-SF	Reference	Reference
SF	3.68 (2.30–5.87)	5.28 (2.99–9.33)
Sensitivity analysis 4		
Non-SF	Reference	Reference
SF	4.00 (2.36–6.78)	5.12 (2.69–9.77)
Sensitivity analysis 5		
Non-SF	Reference	Reference
SF	3.68 (2.30–5.87)	5.17 (2.87–9.29)

Sensitivity analysis 1, HF(+) defined as D-dimer >50 mg/L; sensitivity analysis 2, HF(+) defined as fibrinogen <150 mg/dL; sensitivity analysis 3, using hematoma size as a continuous variable; sensitivity analysis 4, excludes any other AIS of 2 other than head AIS; and sensitivity analysis 5, adding Injury Severity Score as a continuous variable.
OR, odds ratio; aOR, adjusted odds ratio; CI, confidence interval; SF, skull fracture; HF, hyperfibrinolysis; AIS, Abbreviated Injury Scale.
[a]Adjusted for sex, age, Glasgow Coma Scale, type of intracranial injury, and the Traumatic Coma Data Bank classification.

Second, the results of this study, which are robust to multiple sensitivity analyses using a nationwide Japanese multicenter database, are likely to have high internal validity. The results remained robust after adjusting for possible confounding associated with coagulopathy, such as the degrees of intracranial injury and multiple trauma [2,31]. In a previous single-center study we reported that SF, especially depressed SF, were associated with severe HF, but our previous study was small [20]. Therefore, we believe that the current results have high internal validity beyond the limitations of previous studies.

Third, our findings have important clinical implications. When SF is identified on computed tomography (CT), the possibility of HF should be considered, as should additional coagulation testing or periodic coagulation monitoring. In Japan, coagulation testing is

commonly performed in hospitals for patients with TBI, but this practice is not universal. Given that coagulation parameters in TBI can fluctuate over time, follow-up coagulation testing after the initial assessment should be considered in certain cases. Identifying SF, especially vault-SF, on CT might serve as a prompt to assess coagulation status, potentially aiding in the early detection and management of HF. Furthermore, the timing of surgery may require careful consideration in SF cases, particularly in relation to HF. If emergency surgery is required due to signs of impending brain herniation, immediate intervention is unavoidable. However, in cases where there is sufficient time before surgery, stabilizing coagulation status, such as through hemostatic management or transfusion therapy, could be an option. Given the potential for HF in SF patients, preoperative coagulation assessment may provide useful information for surgical planning and perioperative management.

By integrating these considerations into clinical practice, our findings offer insights that could help refine management strategies for TBI with SF.

In this study, higher D-dimer values were found in the SF group. Furthermore, prothrombin time/international normalized ratio and activated partial thromboplastin time were similar between the two groups, but fibrinogen was lower in the SF group compared to non-SF. Since previous studies have reported an inverse relationship between D-dimer and fibrinogen in TBI, the present study also showed a characteristic trend in trauma-induced coagulopathy [32,33]. There was no difference in surgical treatment in HF(+) compared to HF(–), but there were differences in neurological outcomes and in-hospital mortality, suggesting poor prognosis. This study included a larger number of older adults than previous studies conducted in the United States and Europe [34,35]. In Japan, the increase in head trauma among the elderly has become a social problem, and our results reflect the aging of the Japanese society [36]. Furthermore, subgroup analysis by age showed that D-dimer levels were higher in the elderly group than in the nonelderly group, and the frequency of HF(+) was higher, confirming the differences in D-dimer levels by age reported in previous studies, as well as in the present study [26]. Furthermore, the OR of HF(+) was higher in the elderly group than in the nonelderly group.

In general, tissue injury activates the coagulation system in the early stages of trauma by releasing tissue factors and damage-associated molecular patterns (DAMPs); tissue factor complexes with factor VIIa to activate the coagulation system, and DAMPs stimulate inflammatory pathways by releasing multiple mediators. Thrombin production and the formation of fibrin from fibrinogen result in the uncontrolled activation of the coagulation system and the accumulation of large amounts of fibrin, resulting in consumptive coagulopathy [37]. Furthermore, in head trauma, large amounts of tissue plasminogen activator (tPA) are released from the injured endothelial cells, causing so-called primary fibrinolysis that leads to plasminogen activation [38]; HF has also been reported to be associated with increased tPA levels, and is exacerbated by the loss of α2 plasmin inhibitor [39]. After plasminogen activation, plasmin causes cleavage of crosslinked fibrin polymers, leading to formation of fibrin degradation products (secondary fibrinolysis), which in turn causes elevated D-dimer. In this study, D-dimer was used as a surrogate marker for HF. Elevated D-dimer levels, a degradation product of stabilized fibrin that reflects secondary fibrinolysis, may be useful and convenient for representing HF because they result from the degradation of fibrin by plasmin, regardless of a marked increase in free tPA levels or α2-antiplasmin consumption. It should be noted, however, that HF in trauma is a complex mechanism that cannot be expressed by D-dimer alone, as discussed above.

In the present study, we postulated that two mechanisms explain why SF is associated with a high D-dimer, a surrogate finding for HF. First, direct injury from an impact that results in an SF causes severe tissue injury to the brain. The release of tissue factors due to brain tissue injury has been associated with trauma-induced coagulopathy, including HF, and an association between brain tissue injury and high D-dimer levels has also been reported [27]. In fact, as shown in Suppl. 3, findings suggestive of primary brain injuries, such as cerebral contusion, SAH, and IVH, were associated with HF. These findings suggest an association between brain tissue injury and HF. Here, SF reflects the magnitude of the impact to the head and can be considered an objective indicator of impact [18,40]. This impact may result in a degree of brain tissue damage not depicted on CT and thus may have shown an independent association even after adjusting for confounding of intracranial injury. Furthermore, because vault-SF was strongly associated with HF in the subgroup analysis of this study, it may be concluded that the brain area immediately below the tissue injury is larger than the base-SF [41]. Based on these findings, we hypothesized that microscopic tissue injury from the impact causing SF might reflect HF.

Second, we considered the possibility that SF is associated with high D-dimer levels. Previous studies have reported that high D-dimer levels are associated with extremity, rib, and pelvic fractures, and they have shown a trend toward higher levels with larger skeletal injuries [42,43]. Such trends also suggest that they are an indicator of trauma severity [43]. In pediatric head trauma,

D-dimer has received attention as a potential screening tool to avoid CT, based on reports that low D-dimer levels can negate SF [44,45]. Although the detailed mechanism has not been elucidated, it is conceivable that these high D-dimer levels may be caused by endothelial damage to the interplate vessels and periosteal damage due to skull injury, which may be involved in HF. Previous basic studies using a mouse model of TBI have reported that SF induces inflammatory cytokines and that fractures are more likely to induce neuroinflammation than soft tissue injuries [46]. Based on the above, we hypothesized that SF may induce HF, resulting in high D-dimer levels.

We believe these two mechanisms are intertwined based on the available evidence. In particular, the fact that the present study showed a robust association even in analyses adjusted for detailed confounding by the morphology of intracranial injury supports the possibility that the latter, damage to the skull, pericranium, and other structures, may be more related to HF. In summary, our study's findings and proposed mechanisms are interesting and are supported by previous studies, but further research may be necessary to fully understand the pathogenesis of this relationship.

Our study has several limitations. First, we were unable to evaluate HF using detailed coagulation/fibrinolysis parameters. In this study, HF was assessed only by D-dimer and fibrinogen, but in reality, a complex mechanism involving not only secondary fibrinolysis but also tissue-type plasminogen activator release may be involved. In addition, parameters reported in previous studies, such as information on α2-plasmin inhibitor, plasmin-α2-plasmin inhibitor complex, and other parameters obtained by thromboelastometry, were not available for this study [29,47–49]. Although we believe that D-dimer and fibrinogen are important parameters, the results of this study may have been more interesting if such coagulation/fibrinolysis data had been added. Second, while we were able to quantify certain factors, such as hematoma volume, as continuous variables in the sensitivity analysis, SAH and IVH could only be assessed qualitatively due to lack of data. Additionally, detailed intracranial injury assessment using magnetic resonance imaging was not included, making it unclear whether microbleeds or diffuse axonal injury were present. These factors are potential confounders. However, the results of this study remain robust and are unlikely to change. Furthermore, although we analyzed SF by classifying it into base and vault fractures, information on specific fracture patterns, such as depressed or comminuted fractures, was not available. More detailed information on injury morphology could have allowed for further discussion and analysis. Although we attempted to quantify complicated injury morphology as much as possible and to adjust for confounding, including sensitivity analysis, SAH and IVH were not quantified. A detailed intracranial injury assessment with magnetic resonance imaging was not included. Thus, findings of microbleeds or diffuse axonal injury were not clear. These findings would be potential confounders. However, the results of this study are robust and may not change. Third, since the registry did not specify whether D-dimer was collected, the initial practice policy of each hospital likely influenced the data. Excluding patients with missing data may introduce selection bias, such as a skew toward certain institutions or a tendency to include more severe cases.

Conclusions

In this study, we observed an independent association between SF and high D-dimer levels, suggesting an association with HF. We also found that all types of SF were independently associated with an increased risk of HF.

ARTICLE INFORMATION

Author contributions

Conceptualization: GF; Data curation: NS; Formal analysis: GF; Investigation: NS; Writing–original draft: GF; Writing–review & editing: NS. All authors read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

Funding

The authors received no financial support for this study.

Data availability

Data analyzed in this study are available from the Japan Neurotrauma Data Bank. The data are not publicly accessible as they were used under license for this study. However, they are available from the corresponding author upon reasonable request, with permission from the Japan Society of Neurotraumatology.

Supplementary materials

Suppl. 1. Explanation of the Japan Neurotrauma Data Bank 2015.

Suppl. 2. Explanations of Abbreviated Injury Scale, Injury Severity Score, and the Trauma Coma Data Bank classifications.

Suppl. 3. Patients' background divided by HF(+) or HF(–).

Suppl. 4. Crude ORs and adjusted ORs with 95% CIs for each confounding factor (sex, age, GCS, type of intracranial injury,

TCDB classification) for HF.

Suppl. 5. D-dimer values and HF with and without skull fracture in each subgroup divided by age.

Supplementary materials are available from https://doi.org/10.15441/ceem.24.344.

REFERENCES

1. Maegele M. Coagulopathy after traumatic brain injury: incidence, pathogenesis, and treatment options. Transfusion 2013;53 Suppl 1:28S–37S.
2. Samuels JM, Moore EE, Silliman CC, et al. Severe traumatic brain injury is associated with a unique coagulopathy phenotype. J Trauma Acute Care Surg 2019;86:686–93.
3. Gando S, Wada H, Thachil J; Scientific and Standardization Committee on DIC of the International Society on Thrombosis and Haemostasis (ISTH). Differentiating disseminated intravascular coagulation (DIC) with the fibrinolytic phenotype from coagulopathy of trauma and acute coagulopathy of trauma-shock (COT/ACOTS). J Thromb Haemost 2013;11:826–35.
4. Kaufman HH, Hui KS, Mattson JC, et al. Clinicopathological correlations of disseminated intravascular coagulation in patients with head injury. Neurosurgery 1984;15:34–42.
5. Kushimoto S, Kudo D, Kawazoe Y. Acute traumatic coagulopathy and trauma-induced coagulopathy: an overview. J Intensive Care 2017;5:6.
6. Zhang J, Jiang R, Liu L, Watkins T, Zhang F, Dong JF. Traumatic brain injury-associated coagulopathy. J Neurotrauma 2012;29:2597–605.
7. Hayakawa M, Maekawa K, Kushimoto S, et al. High D-dimer levels predict a poor outcome in patients with severe trauma, even with high fibrinogen levels on arrival: a multicenter retrospective study. Shock 2016;45:308–14.
8. Kutcher ME, Cripps MW, McCreery RC, et al. Criteria for empiric treatment of hyperfibrinolysis after trauma. J Trauma Acute Care Surg 2012;73:87–93.
9. Tian HL, Chen H, Wu BS, et al. D-dimer as a predictor of progressive hemorrhagic injury in patients with traumatic brain injury: analysis of 194 cases. Neurosurg Rev 2010;33:359–66.
10. Fujiwara G, Okada Y, Sakakibara T, Yamaki T, Hashimoto N. The association between D-dimer levels and long-term neurological outcomes of patients with traumatic brain injury: an analysis of a nationwide observational neurotrauma database in Japan. Neurocrit Care 2022;36:483–91.
11. Fujiwara G, Murakami M, Maruyama D, Murakami N. Optic nerve sheath diameter as a quantitative parameter associated with the outcome in patients with traumatic brain injury undergoing hematoma removal. Acta Neurochir (Wien) 2023;165:281–7.
12. Zhang J, He M, Song Y, Xu J. Prognostic role of D-dimer level upon admission in patients with traumatic brain injury. Medicine (Baltimore) 2018;97:e11774.
13. Haut RC, Wei F. Biomechanical studies on patterns of cranial bone fracture using the immature porcine model. J Biomech Eng 2017;139.
14. Sahoo D, Deck C, Yoganandan N, Willinger R. Development of skull fracture criterion based on real-world head trauma simulations using finite element head model. J Mech Behav Biomed Mater 2016;57:24–41.
15. Yasui H, Donahue DL, Walsh M, Castellino FJ, Ploplis VA. Early coagulation events induce acute lung injury in a rat model of blunt traumatic brain injury. Am J Physiol Lung Cell Mol Physiol 2016;311:L74–86.
16. Tian Y, Salsbery B, Wang M, et al. Brain-derived microparticles induce systemic coagulation in a murine model of traumatic brain injury. Blood 2015;125:2151–9.
17. Nekludov M, Mobarrez F, Gryth D, Bellander BM, Wallen H. Formation of microparticles in the injured brain of patients with severe isolated traumatic brain injury. J Neurotrauma 2014;31:1927–33.
18. Fujiwara G, Okada Y, Ishii W, et al. Association of skull fracture with in-hospital mortality in severe traumatic brain injury patients. Am J Emerg Med 2021;46:78–83.
19. Berger RP, Fromkin J, Rubin P, Snyder J, Richichi R, Kochanek P. Serum D-dimer concentrations are increased after pediatric traumatic brain injury. J Pediatr 2015;166:383–8.
20. Fujiwara G, Murakami M, Ishii W, et al. Effectiveness of administration of fibrinogen concentrate as prevention of hypofibrinogenemia in patients with traumatic brain injury with a higher risk for severe hyperfibrinolysis: single center before-and-after study. Neurocrit Care 2023;38:640–9.
21. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007;147:573–7.
22. Suehiro E, Fujiyama Y, Koizumi H, Suzuki M. Directions for use of intracranial pressure monitoring in the treatment of severe traumatic brain injury using data from the Japan Neurotrauma Data Bank. J Neurotrauma 2017;34:2230–4.
23. Tokutomi T, Miyagi T, Ogawa T, et al. Age-associated increases in poor outcomes after traumatic brain injury: a report from the Japan Neurotrauma Data Bank. J Neurotrauma 2008;25:1407–14.

24. Maas AI, Hukkelhoven CW, Marshall LF, Steyerberg EW. Prediction of outcome in traumatic brain injury with computed tomographic characteristics: a comparison between the computed tomographic classification and combinations of computed tomographic predictors. Neurosurgery 2005;57:1173–82.
25. Marshall LF, Marshall SB, Klauber MR, et al. The diagnosis of head injury requires a classification based on computed axial tomography. J Neurotrauma 1992;9 Suppl 1:S287–92.
26. Nakae R, Fujiki Y, Takayama Y, et al. Age-related differences in the time course of coagulation and fibrinolytic parameters in patients with traumatic brain injury. Int J Mol Sci 2020;21:5613.
27. Moore EE, Moore HB, Kornblith LZ, et al. Trauma-induced coagulopathy. Nat Rev Dis Primers 2021;7:30.
28. Dong JF, Zhang F, Zhang J. Detecting traumatic brain injury-induced coagulopathy: what we are testing and what we are not. J Trauma Acute Care Surg 2023;94(1S Suppl 1): S50–5.
29. Wada T, Shiraishi A, Gando S, et al. Pathophysiology of coagulopathy induced by traumatic brain injury is identical to that of disseminated intravascular coagulation with hyperfibrinolysis. Front Med (Lausanne) 2021;8:767637.
30. Shibahashi K, Sugiyama K, Okura Y, Hoda H, Hamabe Y. Multicenter retrospective cohort study of "talk and die" after traumatic brain injury. World Neurosurg 2017;107:82–6.
31. Franschman G, Greuters S, Jansen WH, et al. Haemostatic and cranial computed tomography characteristics in patients with acute and delayed coagulopathy after isolated traumatic brain injury. Brain Inj 2012;26:1464–71.
32. Xu DX, Du WT, Li X, Wu ZX, Yu GF. D-dimer/fibrinogen ratio for the prediction of progressive hemorrhagic injury after traumatic brain injury. Clin Chim Acta 2020;507:143–8.
33. Nakae R, Yokobori S, Takayama Y, et al. A retrospective study of the effect of fibrinogen levels during fresh frozen plasma transfusion in patients with traumatic brain injury. Acta Neurochir (Wien) 2019;161:1943–53.
34. Steyerberg EW, Wiegers E, Sewalt C, et al. Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. Lancet Neurol 2019;18:923–34.
35. McCrea MA, Giacino JT, Barber J, et al. Functional outcomes over the first year after moderate to severe traumatic brain injury in the prospective, longitudinal TRACK-TBI study. JAMA Neurol 2021;78:982–92.
36. Yokobori S, Saito K, Sasaki K, et al. Treatment of geriatric traumatic brain injury: a nationwide cohort study. J Nippon Med Sch 2021;88:194–203.
37. Gando S. Tissue factor in trauma and organ dysfunction. Semin Thromb Hemost 2006;32:48–53.
38. Chapman MP, Moore EE, Moore HB, et al. Overwhelming tPA release, not PAI-1 degradation, is responsible for hyperfibrinolysis in severely injured trauma patients. J Trauma Acute Care Surg 2016;80:16–25.
39. Barrett CD, Moore HB, Vigneshwar N, et al. Plasmin thrombelastography rapidly identifies trauma patients at risk for massive transfusion, mortality, and hyperfibrinolysis: a diagnostic tool to resolve an international debate on tranexamic acid? J Trauma Acute Care Surg 2020;89:991–8.
40. Fujiwara G, Okada Y, Shiomi N, Sakakibara T, Yamaki T, Hashimoto N. Derivation of coagulation phenotypes and the association with prognosis in traumatic brain injury: a cluster analysis of nationwide multicenter study. Neurocrit Care 2024;40:292–302.
41. Boldsen JL, Milner GR, Weise S. Cranial vault trauma and selective mortality in medieval to early modern Denmark. Proc Natl Acad Sci U S A 2015;112:1721–6.
42. Liu C, Song Y, Zhao J, et al. Elevated D-dimer and fibrinogen levels in serum of preoperative bone fracture patients. Springerplus 2016;5:161.
43. Johna S, Cemaj S, O'Callaghan T, Catalano R. Effect of tissue injury on D-Dimer levels: a prospective study in trauma patients. Med Sci Monit 2002;8:CR5–8.
44. Nozawa M, Mishina H, Tsuji S, Takayama JI. Low plasma D-dimer predicts absence of intracranial injury and skull fracture. Pediatr Int 2020;62:22–8.
45. Langness S, Ward E, Halbach J, et al. Plasma D-dimer safely reduces unnecessary CT scans obtained in the evaluation of pediatric head trauma. J Pediatr Surg 2018;53:752–7.
46. Zvejniece L, Stelfa G, Vavers E, et al. Skull fractures induce neuroinflammation and worsen outcomes after closed head injury in mice. J Neurotrauma 2020;37:295–304.
47. Bugaev N, Como JJ, Golani G, et al. Thromboelastography and rotational thromboelastometry in bleeding patients with coagulopathy: practice management guideline from the Eastern Association for the Surgery of Trauma. J Trauma Acute Care Surg 2020;89:999–1017.
48. Baksaas-Aasen K, Gall LS, Stensballe J, et al. Viscoelastic haemostatic assay augmented protocols for major trauma haemorrhage (ITACTIC): a randomized, controlled trial. Intensive Care Med 2021;47:49–59.
49. Takahashi H, Urano T, Takada Y, Nagai N, Takada A. Fibrinolytic parameters as an admission prognostic marker of head injury in patients who talk and deteriorate. J Neurosurg 1997;86:768–72.

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):74-80
https://doi.org/10.15441/ceem.25.036

Original Article

pISSN: 3022-1811
eISSN: 2383-4625

A pilot study of recurrent abdominal pain in the emergency department: low-risk disease associated with high pain severity and frequent opioid administration

Soroush Shahamatdar[1], Felice Yang[2], Meylakh Barshay[3], Ryan Heidish[2], Aditya Loganathan[2], Taylor Bolden[2], William Huang[2], Andrew C. Meltzer[2]

[1]David Geffen School of Medicine at UCLA, Los Angeles, CA, USA
[2]Department of Emergency Medicine, The George Washington University School of Medicine and Health Sciences, Washington, DC, USA
[3]Warren Alpert Medical School of Brown University, Providence, RI, USA

Objective Abdominal pain is the most common emergency department (ED) chief complaint, with many patients experiencing recurrent episodes due to non–life-threatening etiologies such as disorders of gut-brain interaction. This pilot study aimed to characterize patients with recurrent low-risk abdominal pain, focusing on pain severity, management, biopsychosocial factors, opioid use, and 30-day return visits.

Methods This prospective, observational pilot study enrolled adult ED patients with recurrent abdominal pain at a single academic center between July 2022 and June 2023. Inclusion required at least one similar episode in the prior year with symptom resolution between episodes. Exclusions included unstable clinical status or high-risk conditions. Patient-reported outcomes, social determinants of health, and clinical data were collected. Primary outcomes included pain severity, opioid use, and 30-day return visit rates.

Results A total of 101 participants were enrolled (mean age, 43.7 years; 65.3% female; 69.3% Black). Pain severity was high (triage pain score, 7.1±2.6). Frequent prior computed tomography imaging was noted in 56.4% of participants. Opioids were administered in 50 participants (49.5%), while Patient-Reported Outcomes Measurement Information System (PROMIS)-29 scores highlighted risks of anxiety (T-score, 56.0±11.1) and pain interference (T-score, 60.8±8.2). Return visits occurred in 11 participants (10.9%) within 30 days.

Conclusion In this pilot study, patients with recurrent low-risk abdominal pain showed high symptom burden and healthcare utilization. Targeted interventions addressing biopsychosocial factors and improving pain management are needed to reduce ED revisits and improve outcomes.

Keywords Abdominal pain; Emergency department; Opioid use; Social determinants

Received: February 8, 2025
Revised: June 12, 2025
Accepted: June 16, 2025

Correspondence to:
Andrew C. Meltzer
(ameltzer@mfa.gwu.edu)

How to cite this article:
Shahamatdar S, Yang F, Barshay M, Heidish R, Loganathan A, Bolden T, Huang W, Meltzer AC. A pilot study of recurrent abdominal pain in the emergency department: low-risk disease associated with high pain severity and frequent opioid administration.
Clin Exp Emerg Med 2026;13(1):74-80.
https://doi.org/10.15441/ceem.25.036

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

© 2026 The Korean Society of Emergency Medicine

Capsule Summary

What is already known

Abdominal pain is the most common emergency department chief complaint, with many patients experiencing recurrent episodes due to non–life-threatening etiologies such as disorders of gut-brain interaction.

What is new in the current study

This study investigates the characteristics and outcomes of patients presenting with recurrent low-risk abdominal pain with a focus on pain severity, biopsychosocial factors, opioid use, and 30-day return visits. Our findings highlight the high symptom burden and frequent healthcare utilization among this patient population, as well as gaps in pain management and communication regarding the cause of symptoms. The study underscores the need for tailored interventions that address biopsychosocial contributors in order to improve management strategies in the emergency department.

INTRODUCTION

Abdominal pain is the most common reason for emergency department (ED) visits in the United States, accounting for approximately 7 million visits per year [1]. For many patients, abdominal pain is not a one-time event but a recurrent symptom with repeat episodes and multiple ED visits. Approximately 15% to 34% of patients with recurrent abdominal pain qualify as super-utilizers, defined as having four or more ED visits per year [2]. The causes of recurrent abdominal pain are highly heterogeneous including structural disease such as chronic pancreatitis, diverticulitis, and inflammatory bowel disease and nonstructural causes such as disorders of gut-brain interaction [3–5]. Patients with recurrent abdominal pain often undergo multiple diagnostic studies such as laboratory tests, computed tomography (CT) scans and ultrasounds. As a result, visits for ED patients with abdominal pain are resource-intensive and costly [6]. In addition, 44% to 59% of ED abdominal pain patients are treated with opioid analgesics, which are associated with adverse events [7]. Patients presenting with low-risk abdominal pain often exhibit symptoms that closely resemble more acute, high-risk conditions, resulting in similar diagnostic workups that expose them to unnecessary iatrogenic radiation, extensive laboratory testing, significant healthcare costs, and prescription of opioid medications, which can contribute to dependence, adverse side effects, and long-term misuse [8,9]. These patients could be more effectively managed in an outpatient setting using a biopsychosocial approach that emphasizes longitudinal care, addresses underlying psychosocial contributors, and promotes lifestyle modification, all while reducing the strain on EDs and conserving critical acute care resources. Given the significant patient stress, cost, and overall burden on the healthcare system, patients with recurring episodes of abdominal pain highlight a gap in the healthcare system, contributing to unmet patient needs. The objective of this study was to characterize patients with recurrent abdominal pain and describe their symptoms, the use of opioid analgesia, the efficacy of pain control, and the frequency of 30-day ED return visits.

METHODS

Ethics statement

This study was approved by the Institutional Review Board of The George Washington University (No. NCR213728). Informed consent was obtained from all patients.

Study setting and design

This prospective case-cohort observational study was conducted at The George Washington University Hospital in Washington, DC, USA, from July 4, 2022, to June 1, 2023. Patients were screened for recurrent low-risk abdominal pain, excluding patients with established high-risk characteristics. Study participants completed an electronic survey to evaluate symptomatology, comorbidities, and psychosocial risk factors. In addition, an automated report generated from the electronic health record after the index visit was reviewed to assess 30-day outcomes.

Study population

Adult patients (≥18 years old) presenting to the ED with a chief complaint of abdominal pain and at least one reported episode of similar pain in the past year were screened for this study. Recurrence was defined as patients with a similar episode of abdominal pain at least 30 days before the current ED visit with a period of resolution between episodes such that pain was not continuous, i.e., episodic pain. Patients were excluded if they met any of the following criteria: altered mental status, prisoner/ward of state, pregnant, or non–English-speaking or did not meet all low-risk criteria. Low-risk criteria were established by the GRACE-2 (Guidelines for Reasonable and Appropriate Care in the Emergen-

cy Department 2) writing group [1] and include the following: (1) stable vital signs; (2) age range of 18 to 65 years; (3) not pregnant; (4) no history or physical findings of acute abdominal pathology; (5) no acute trauma within the last 7 days; (6) no prior organ transplantation or immunosuppression; (7) no abdominal surgery within 30 days and/or active cancer; (8) no history of inflammatory bowel disease; (9) no previous bowel obstructions; and (10) no active severe psychiatric illnesses requiring urgent stabilization.

The cohort was a convenience sample enrolled by research staff who screened ED patients on weekdays from 9 AM to 7 PM. Once enrolled, participants were evaluated to assess for pain severity, frequency of symptoms, and confidence in diagnosis. Other validated scoring tools were used to assess for substance use risk (Alcohol, Smoking and Substance Involvement Screening Test, ASSIST) [10], social determinants of health (Accountable Health Communities [AHC] health-related social needs [HRSN]) [11], and health-related quality of life (Patient-Reported Outcomes Measurement Information System [PROMIS]-29) [12]. Additional data on clinical outcomes, including ED management, ED disposition, ED diagnosis, and 30-day ED revisits, were collected through automated electronic health record reports. The primary ICD-10 (International Classification of Diseases, 10th Revision) diagnosis code associated with this visit was also recorded.

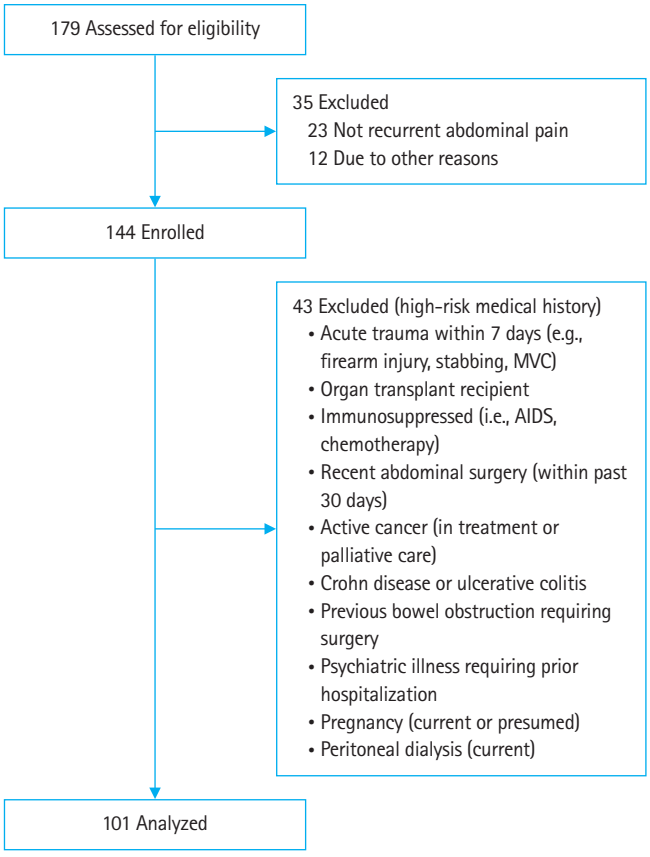


Fig. 1. Study flowchart. MVC, motor vehicle collision.

Analysis

Descriptive statistics were used to summarize demographic, clinical, and patient-reported outcome data. Continuous variables, such as pain scores (10-point numerical rating system) and PROMIS-29 T-scores, were reported as mean ± standard deviation, while categorical variables, including sex, race, and opioid administration, were presented as frequencies and percentages. Associations between 30-day return visits to the ED and patient-reported outcomes were assessed using a multivariate logistic regression. The model was also adjusted for age and sex. Unadjusted and adjusted odds ratios were reviewed, as well as 95% confidence intervals (CIs). All analyses were conducted using SAS ver. 9.4 (SAS Institute Inc), with a two-tailed significance threshold of $P < 0.05$.

RESULTS

As shown in Fig. 1, a total of 101 participants were included in the study with a mean age of 43.7 ± 16.7 years. Sex distribution was predominantly female, with 66 participants (65.3%) identified as women and 34 participants (33.7%) as men. Regarding racial distribution, 70 participants (69.3%) reported being Black. Participants reported high levels of pain severity with pain on multiple days. The mean triage pain score was 7.1 ± 2.6 and the mean pain score at time of discharge was 5.0 ± 3.4 on a 10-point scale. Approximately half of the participants (n = 50, 49.5%) were treated with opioids in the ED, while 13.9% of participants were prescribed opioids from the ED (Table 1). When asked about their understanding of their pain, 37 participants (36.6%) reported being informed of the cause by a provider, while 22 (21.8%) indicated they were told but did not understand. A notable 26 participants (25.7%) stated they were never informed of the cause, while 2 (2.0%) were uncertain if they had been told. Additionally, 14 participants (13.9%) reported never seeking medical attention for their pain. Prior CT imaging was also noted, with 57 (56.4%) having received at least one CT scan in the past 5 years for this abdominal pain.

Table 2 provides insights into questionnaire scores among participants, focusing on metrics such as physical function and anxiety measured using the PROMIS-29 tool. Of the 101 participants who completed the assessments, the mean T-score for physical

Table 1. Demographic variables among participants (n=101)

Characteristic	Value
Age (yr)	43.7 ± 16.7
Sex[a]	
Male	34 (33.7)
Female	66 (65.3)
Other (transgender, nonbinary)	1 (1.0)
Race[a]	
Black	70 (69.3)
White	21 (20.8)
Asian	2 (2.0)
Other	4 (4.0)
More than one race	1 (1.0)
Prefer not to answer	3 (3.0)
Ethnicity[a]	
Hispanic or Latino	5 (5.0)
Not Hispanic or Latino	91 (90.1)
Prefer not to answer	5 (5.0)
Insurance status	
Medicaid	34 (33.7)
Medicare	17 (16.8)
Private	25 (24.8)
Uninsured	12 (11.9)
Other	13 (12.9)
Past medical history	
High blood pressure	38 (37.6)
Diabetes	18 (17.8)
Congestive heart disease	5 (5.0)
COPD	6 (5.9)
Coronary artery disease	3 (3.0)
Depression	26 (25.7)
Anxiety	30 (29.7)
Pain score (range, 0–10)	
Triage pain score	7.1 ± 2.6
Current pain score[a]	6.1 ± 2.9
Discharge pain score	5.0 ± 3.4
Questionnaire	
In the past 7 days, how often did you have abdominal pain?	
More than once a day	30 (29.7)
Once a day	17 (16.8)
At least 2–6 days	36 (35.6)
One day only	14 (13.9)
Never	3 (3.0)
Missing data	1 (1.0)
In the past 7 days, how much did your abdominal pain interfere with your day-to-day activities?	
Quite a bit, very much	67 (66.3)
Not at all, a little bit, or somewhat	33 (32.7)
Missing data	1 (1.0)
Prior episode of similar symptoms?	
1 Week prior	20 (19.8)
More than 1 week and less than a month	20 (19.8)
Between 1 month and 6 months prior	33 (32.7)
Between 6 months and 1 year prior	14 (13.9)
More than 1 year prior	10 (9.9)
Not sure or unable to answer	3 (3.0)
Missing data	1 (1.0)
≥ 1 CT scan in the past 5 years for this abdominal pain	57 (56.4)
Has a doctor or medical provider been able to tell you what was causing the pain?	
Yes, I was told what was wrong and the cause of my pain is clear	37 (36.6)
I was told the cause of my pain, but I do not understand	22 (21.8)
No, I was never told the cause	26 (25.7)
Uncertain if I was told the cause	2 (2.0)
I never saw a doctor	14 (13.9)
Administered medication	
Opioid in the ED	50 (49.5)
NSAID[a] in the ED	21 (20.8)
APAP[a] in the ED	39 (38.6)
Alternative analgesia in the ED	12 (11.9)
ED course	
CT scan in the ED	61 (60.4)
Ultrasound in the ED	20 (19.8)
Discharged to home	68 (67.3)
Prescribed opioid analgesia at discharge	14 (13.9)

Values are reported as mean±standard deviation or number (%).
COPD, chronic obstructive pulmonary disease; CT, computed tomography; ED, emergency department; NSAID, nonsteroidal anti-inflammatory drug; APAP, acetaminophen.
[a]Self-reported.

function was 43.8 ± 10.2, indicating below-average physical functioning in the sample. Anxiety levels were elevated, with a mean T-score of 56.0 ± 11.1, suggesting significant psychological distress. For all participants, 30-day outcomes were reviewed to assess return visits and ED course. Of the participants, 90 (89.1%) did not return to the ED during the 30-day follow-up period, while 7 participants (6.9%) returned once, and 4 participants (4.0%) returned twice.

Further analysis examined associations between return visits and clinical or demographic factors. This study found no statistically significant associations between patient-reported factors and repeat ED visits within 30 days. While trends suggest that high-impact chronic pain and financial strain may increase the likelihood of return visits, the wide CIs indicated uncertainty (Table 3). Initial findings suggest that individuals with more frequent ED visits may have higher levels of unresolved pain or chronic conditions contributing to repeat healthcare utilization. This distribution underscores the importance of targeted follow-up and interventions to reduce frequent ED utilization among at-risk populations.

DISCUSSION

Abdominal pain is one of the most common chief complaints to US EDs and represents a heterogeneous group of diseases with a

Table 2. Questionnaire scores among participants (n=101)

Questionnaire	Value
PROMIS-29 score[a]	
Physical function T-score	43.8 ± 10.2
Anxiety T-score	56.0 ± 11.1
Depression T-score	51.7 ± 10.5
Fatigue T-score	54.3 ± 10.4
Sleep disturbance T-score	55.5 ± 9.9
Ability to participate in social activities T-score	52.2 ± 9.6
Pain interference T-score	60.8 ± 8.2
Global pain score[b]	7.0 ± 2.4
Moderate- to high-risk substance use (ASSIST score)	
Alcohol (n = 61)	9.2 ± 8.7
Tobacco (n = 31)	14.9 ± 9.0
Cannabis (n = 33)	14.1 ± 9.4
Cocaine (n = 4)	14.5 ± 10.4
Amphetamine (n = 3)	16.3 ± 10.6
Sedative (n = 2)	12.0 ± 0.0
Hallucinogen (n = 6)	12.3 ± 9.9
Opioid (n = 5)	11.6 ± 8.3
Simplified chronic pain scale	
Grade 0 (chronic pain absent)	68 (67.3)
Grade 1 (mild or bothersome chronic pain)	17 (16.8)
Grade 2 (high-impact chronic pain)	16 (15.8)
Unmet health-related social strain (AHC-HRSN score)	
Financial strain	37 (36.6)
Employment	15 (14.9)
Family and community support	23 (22.8)
Education	21 (20.8)
Physical activity	59 (58.4)

Values are presented as mean±standard deviation or number (%).
PROMIS, Patient-Reported Outcomes Measurement Information System; ASSIST, Alcohol, Smoking and Substance Involvement Screening Test; AHC, Accountable Health Communities; HRSN, health-related social needs.
[a]A higher score indicates more adverse outcomes in the listed category, aside from physical function, where lower scores indicate adverse outcomes. [b]Global pain score is a 0–10 pain scale representing the patient's average pain in the last week.

Table 3. Association between repeat ED visit within 30 days of index visit and patient-reported factors

Factor	OR (95% CI)	aOR (95% CI)[a]
PROMIS-29 score		
Physical function T-score	0.97 (0.92–1.04)	0.99 (0.92–1.06)
Anxiety T-score	1.06 (0.99–1.13)	1.06 (0.99–1.14)
Depression T-score	1.05 (0.99–1.11)	1.05 (0.99–1.12)
Fatigue T-score	1.02 (0.96–1.08)	1.04 (0.97–1.11)
Sleep disturbance T-score	0.99 (0.93–1.06)	1.00 (0.94–1.07)
Ability to participate in social activities T-score	0.95 (0.89–1.02)	0.95 (0.88–1.02)
Pain interference T-score	1.02 (0.94–1.10)	1.03 (0.94–1.12)
Global pain score	1.36 (0.96–1.93)	1.34 (0.95–1.90)
ASSIST score		
Alcohol	0.98 (0.85–1.14)	1.02 (0.88–1.18)
Cannabis	1.04 (0.91–1.19)	1.70 (0.21–13.82)
Pain score at discharge	1.01 (0.84–1.21)	0.99 (0.81–1.20)
Grade 0 (chronic pain absent)	1 (Reference)	1 (Reference)
Grade 1 (mild or bothersome chronic pain)	0.55 (0.06–4.75)	0.60 (0.07–5.49)
Grade 2 (high-impact chronic pain)	2.01 (0.46–8.83)	2.02 (0.43–9.57)
AHC-HRSN score		
Financial strain	1.49 (0.42–5.25)	1.66 (0.45–6.18)
Employment	0.54 (0.06–4.52)	0.71 (0.07–6.86)
Family and community support	0.72 (0.14–3.60)	0.65 (0.12–3.41)
Education	2.42 (0.64–9.22)	2.05 (0.51–8.27)
Physical activity	0.82 (0.23–2.87)	0.70 (0.18–2.75)

ED, emergency department; OR, odds ratio; CI, confidence interval; aOR, adjusted odds ratio; PROMIS, Patient-Reported Outcomes Measurement Information System; ASSIST, Alcohol, Smoking and Substance Involvement Screening Test; AHC, Accountable Health Communities; HRSN, health-related social needs.
[a]Adjusted by sex and age.

wide variety of treatments, diagnostic tests, and pain management. Patients with recurrent abdominal pain may not be offered treatment due to uncertainty of diagnosis and concerns about the dangers of opioid analgesia. In particular, patients presenting with low-risk recurrent abdominal pain often exhibit symptoms that closely mimic those of high-risk, acute conditions, leading to substantial resource utilization, including frequent imaging and emergency evaluations, despite a low likelihood of serious pathology; these patients may be more appropriately managed in the outpatient setting, where primary care providers and gastroenterologists can focus on developing long-term care plans, addressing lifestyle factors, and assessing comorbid conditions in a more sustainable and patient-centered manner. We chose to study patients presenting with low-risk abdominal pain to better characterize this patient population and validate the GRACE-2 guidelines for low-risk abdominal pain. In this study, patients with recurrent abdominal pain showed minimal improvement in pain during their ED stay. In addition, they reported high levels of pain interference with activities of daily living and often received opioids in the ED. This pattern is inconsistent with recommendations to limit opioid treatment to the lowest effective dose, and increasing dose based on severity of pain and treatment side effects for chronic noncancer pain [13]. The high rates of opioid treatment observed among patients with low-risk recurrent abdominal pain suggest that providers may not be consistently prioritizing strategies to minimize the risk of opioid dependence; given the frequency with which these patients present to the ED, this practice places a large population at unnecessary risk for opioid-related harm despite the availability of safer, more sustainable management strategies outside the acute care setting. While trends suggest that high-impact chronic pain and financial strain may increase the likelihood of return visits, further research with larger sample sizes is needed to clarify these potential asso-

ciations.

It is possible that improving diagnosis will lead to more effective care, reduced costs from repeat ED visits, and improved health outcomes for abdominal pain patients. The need for answers is one of the major motivations for patients to present to the ED. Proper education regarding a diagnosis has been shown to have a therapeutic effect [14]. In this study, recurrent abdominal pain patients do not appear to have received effective communication regarding the cause of their pain. In comparison, those with a more serious diagnosis may be more likely to claim that they were told "what was wrong, and the cause of their pain was clear" by a doctor or medical provider.

Managing chronic abdominal pain can require a multifaceted approach. A tailored approach that does not involve an increase in opioid analgesia is preferred. Current guidelines on the management of chronic abdominal pain include initially treating with proton pump inhibitors, first-line dietary treatments, antidiarrheals, and laxatives before turning to opioid analgesics [15]. Such treatments may serve to improve pain score outcomes while avoiding the potential side effects of opioid analgesia. In addition, psychiatric treatments such as cognitive behavioral therapies and clinical hypnosis have been shown effective in disorders of gut-brain interaction like irritable bowel syndrome [16]. Integrating psychotherapy in the ED alongside existing medical treatments could also be beneficial in lowering pain scores of patients with benign diagnoses. Additionally, psychiatric comorbidities have been found to be more common in patients who are frequent utilizers of the ED [17]. A biopsychosocial approach to pain management has the potential to be beneficial for reducing pain scores and preserving ED resources.

Although socioeconomic factors were not significantly associated with repeat ED visits, trends suggest financial and social determinants may influence healthcare utilization. Financial strain was potentially linked to increased ED reliance, reflecting barriers to primary care and challenges managing chronic conditions. Prior studies have shown similar findings, identifying financial strain as being associated with the patient identifying the ED as their primary resource for obtaining care, while reporting adverse social determinants of health was correlated with actual ED visits [18]. Employment and education showed no clear associations, although work constraints might limit outpatient follow-up, and higher education levels could lead to proactive care-seeking behavior. Physical activity showed no significant impact but remains important for overall health. These findings highlight the need for larger, more longitudinal studies to clarify these relationships. Addressing financial barriers, improving outpatient access, and strengthening social support systems may help reduce avoidable ED visits and improve outcomes.

Limitations

The limitations of this study should be acknowledged. First, the single-center design conducted at an academic ED may limit the generalizability of findings to other settings or patient populations with differing demographic or clinical characteristics. Second, the use of convenience sampling during limited hours introduces selection bias, potentially excluding patients who present outside of these timeframes. Third, as a pilot study, the relatively small sample size (n = 101) limits the statistical power to detect associations and draw definitive conclusions. Additionally, reliance on self-reported data for variables such as pain interference and psychosocial factors introduces the potential for recall and social desirability biases. The exclusion of non-English-speaking patients further contributes to selection bias and restricts the applicability of findings to linguistically diverse populations. Moreover, the study's focus on outcomes within a 30-day follow-up period provides a narrow time frame that does not capture longer-term healthcare utilization or patient outcomes. Despite adjustments for age and sex, the potential for residual confounding remains, as unmeasured factors may influence the observed associations between patient-reported outcomes and ED return visits. Finally, the specificity of the inclusion criteria, which exclude patients with high-risk conditions, limits the applicability of these findings to broader populations with recurrent abdominal pain, including those with higher-risk diagnoses. Addressing these limitations in future studies will be critical to enhancing the understanding and management of recurrent abdominal pain in ED settings.

Conclusions

This pilot study emphasizes the challenges of managing recurrent abdominal pain in ED patients, revealing that those with benign diagnoses often leave the ED unsure about the cause of their pain and with less effective pain management compared to patients with serious diagnoses. These findings highlight the need for a more nuanced approach to the diagnosis and treatment of patients with recurrent low-risk abdominal pain, emphasizing the need for tailored treatment strategies that supplement analgesic approaches to pain management. Future research should investigate biopsychosocial factors to enhance therapeutic interactions and improve overall management of low-risk abdominal pain in the ED setting.

ARTICLE INFORMATION

Author contributions

Conceptualization: ACM; Data curation: ACM; Formal analysis: all authors; Investigation: ACM; Methodology: FY, MB; Writing–original draft: all authors; Writing–review & editing: all authors. All authors read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

Funding

The authors received no financial support for this study.

Data availability

Data analyzed in this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Broder JS, Oliveira J E Silva L, Bellolio F, et al. Guidelines for Reasonable and Appropriate Care in the Emergency Department 2 (GRACE-2): low-risk, recurrent abdominal pain in the emergency department. Acad Emerg Med 2022;29:526–60.
2. Carpenter CR, Griffey RT, Mills A, et al. Repeat computed tomography in recurrent abdominal pain: an evidence synthesis for guidelines for reasonable and appropriate care in the emergency department. Acad Emerg Med 2022;29:630–48.
3. Rome Foundation. Rome IV criteria [Internet]. Rome Foundation; [cited 2021 Jan 16]. Available from: https://theromefoundation.org/rome-iv/rome-iv-criteria/
4. Cleveland Clinic. Gastrointestinal diseases [Internet]. Cleveland Clinic; [cited 2021 Jan 14]. Available from: https://my.clevelandclinic.org/health/articles/7040-gastrointestinal-diseases
5. Drossman DA, Hasler WL. Rome IV: functional GI disorders: disorders of gut-brain interaction. Gastroenterology 2016;150:1257–61.
6. Brenner DM, Brandt LJ, Fenster M, et al. Rare, overlooked, or underappreciated causes of recurrent abdominal pain: a primer for gastroenterologists. Clin Gastroenterol Hepatol 2023;21:264–79.
7. Miller AC, Khan AM, Castro Bigalli AA, et al. Neuroleptanalgesia for acute abdominal pain: a systematic review. J Pain Res 2019;12:787–801.
8. Daniels J, Griffiths M, Fisher E. Assessment and management of recurrent abdominal pain in the emergency department. Emerg Med J 2020;37:515–21.
9. Graff LG 4th, Robinson D. Abdominal pain and emergency department evaluation. Emerg Med Clin North Am 2001;19:123–36.
10. Humeniuk R, Henry-Edwards S, Ali R, Poznyak V, Monteiro MG; World Health Organization. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): manual for use in primary care. World Health Organization; 2010.
11. Billioux A, Verlander K, Anthony S, Alley D. Standardized screening for health-related social needs in clinical settings: the Accountable Health Communities screening tool [Internet]. NAM Perspectives; 2017 [cited 2025 Mar 1]. Available from: https://doi.org/10.31478/201705b
12. Luijten MA, Haverman L, van Litsenburg RR, Roorda LD, Grootenhuis MA, Terwee CB. Advances in measuring pediatric overall health: the PROMIS® Pediatric Global Health scale (PGH-7). Eur J Pediatr 2022;181:2117–25.
13. Fine PG, Mahajan G, McPherson ML. Long-acting opioids and short-acting opioids: appropriate use in chronic pain management. Pain Med 2009;10 Suppl 2:S79–88.
14. Correia JC, Waqas A, Assal JP, et al. Effectiveness of therapeutic patient education interventions for chronic diseases: a systematic review and meta-analyses of randomized controlled trials. Front Med (Lausanne) 2023;9:996528.
15. Keefer L, Ko CW, Ford AC. AGA clinical practice update on management of chronic gastrointestinal pain in disorders of gut-brain interaction: expert review. Clin Gastroenterol Hepatol 2021;19:2481–8.
16. Atkins M, Burton Murray H, Staller K. Assessment and management of disorders of gut-brain interaction in patients with eating disorders. J Eat Disord 2023;11:20.
17. Ondler C, Hegde GG, Carlson JN. Resource utilization and health care charges associated with the most frequent ED users. Am J Emerg Med 2014;32:1215–9.
18. Davis CI, Montgomery AE, Dichter ME, Taylor LD, Blosnich JR. Social determinants and emergency department utilization: findings from the Veterans Health Administration. Am J Emerg Med 2020;38:1904–9.

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):81-85
https://doi.org/10.15441/ceem.24.321

Brief Research Report

pISSN: 3022-1811
eISSN: 2383-4625

Performance of a transesophageal echocardiography probe at temperature monitoring during simulated hypothermia and rewarming

Madeline Larochelle[1], Margarita Popova[1], David Mackenzie[1,2], Andrew Fried[1,2], Peter Croft[1,2], Joshua Rehberg[1,2], Christina Wilson[1,2]

[1]Department of Emergency Medicine, MaineHealth Maine Medical Center, Portland, ME, USA
[2]Tufts University School of Medicine, Boston, MA, USA

Received: September 24, 2024
Revised: December 17, 2024
Accepted: December 18, 2024

Correspondence to:
Christina Wilson
(christina.wilson@mainehealth.org)

Objective To determine whether a transesophageal echocardiography (TEE) probe can accurately measure temperature and be used to monitor temperature changes over time without overheating in an experimental model of hypothermia and rewarming.

Methods A 6-L water bath was heated with a sous vide immersion circulator to 24, 28, 32, and 36 °C to simulate severe hypothermia, moderate hypothermia, mild hypothermia, and normothermia, respectively. A TEE probe, esophageal temperature probe, and bladder temperature probe were used to measure temperature every 60 seconds for 15 minutes.

Results The TEE probe reported temperatures with a mean difference of 0.60 °C (95% confidence interval [CI], 0.51 to 0.69 °C) compared to the sous vide immersion circulator. The esophageal probe and bladder probe reported temperatures with a mean difference of −0.19 °C (95% CI, −0.23 to −0.14 °C) and −0.20 °C (95% CI, −0.26 to −0.14 °C), respectively.

Conclusion During this simulation, the TEE tip temperature did not increase beyond the expected changes produced by water temperature. The probe temperature was less accurate than the esophageal and bladder temperature probes but demonstrated precision in monitoring temperature changes and stable hypothermia. Based on this study, TEE probes should not be relied upon for an accurate initial temperature but can likely be used to monitor changes in temperature over time.

Keywords Transesophageal echocardiography; Resuscitation; Hypothermia; Cardiac arrest

Capsule Summary

What is already known

Transesophageal echocardiography is increasingly being used to guide resuscitation during cardiac arrest in the emergency department, including in patients with hypothermic cardiac arrest.

What is new in the current study

The transesophageal echocardiography probe may be useful for tracking temperature changes during rewarming efforts and does not heat up with simulated use.

How to cite this article:
Larochelle M, Popova M, Mackenzie D, Fried A, Croft P, Rehberg J, Wilson C. Performance of a transesophageal echocardiography probe at temperature monitoring during simulated hypothermia and rewarming. Clin Exp Emerg Med 2026;13(1):81-85. https://doi.org/10.15441/ceem.24.321

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

© 2026 The Korean Society of Emergency Medicine

INTRODUCTION

Temperature monitoring during resuscitation from cardiac arrest during accidental hypothermia is essential to monitor the response to rewarming efforts and to guide further resuscitation [1]. Rectal, bladder, and esophageal temperature probes can provide continuous temperature measurement but have potential limitations. Active rewarming with peritoneal lavage or warm water instillation into the bladder can artificially elevate rectal and bladder temperature readings [2]. Esophageal temperature measurement is often considered superior to rectal and bladder temperature measurements given its proximity to the heart and major blood vessels and has been shown to respond more quickly to dynamic changes in core temperature [3]. Transesophageal echocardiography (TEE) is increasingly being used to guide resuscitation during cardiac arrest [4]. TEE probes routinely measure and display the probe tip temperature during use and could potentially be used to monitor temperature during hypothermic cardiac arrest as an adjunct or alternative to other temperature sensors. Simultaneous cardiac ultrasound imaging and temperature monitoring with the TEE probe could streamline resuscitation efforts by reducing the need for additional interventions by eliminating the need for an additional temperature monitoring probe.

Ultrasound probe manufacturers warn that operators should expect heating of the TEE probe with routine use. Continued operation in two-dimensional (2D) mode results in the lowest transducer surface temperatures, and using a still image can reduce temperatures further. In manufacturer studies, there was a rise in TEE tip temperature of 4.9 °C with simulated use and of 10.1 °C at the controls and settings anticipated to produce the maximum temperature [5]. A study in patients undergoing cardiac surgery with cardiopulmonary bypass showed that temperature monitoring with a TEE probe during periods of cooling and rewarming was superior to monitoring with a nasopharyngeal probe without any report of probe overheating [6].

This study uses an experimental model of hypothermia and rewarming to determine if a TEE probe can accurately measure temperature and be used to monitor temperature changes over time without overheating. In a prior study, temperatures were measured during cardiopulmonary bypass and only reached mild hypothermia; we aim to evaluate TEE temperature performance in severe, moderate, and mild hypothermia ranges [6].

METHODS

Study design

A 6-L water bath was heated to a starting temperature of 24 °C using a sous vide immersion circulator (Anova Precision Cooker 3.0, Anova Applied Electronics Inc), a heating device that circulates water to maintain a uniform water temperature. The starting temperature of 24 °C was selected to simulate severe hypothermia [1]. Once the water bath was at temperature, as determined by the sous vide immersion circulator, a TEE probe (SonoSite TEEx 8-3 MHz Transducer, Fujifilm SonoSite Inc), an esophageal temperature probe (9F General Purpose Temperature 400 Series, DeRoyal Industries Inc), and a bladder temperature probe (14F Surestep Temperature Sensing Foley System, BD) were added to the water bath (Fig. 1). The tip of each device was placed in approximately the same location and depth in the water bath and was monitored throughout the experiment, making adjustments as needed. The water bath was maintained at 24 °C for 15 minutes, and the temperature reading of each device was recorded every 60 seconds. After 15 minutes, the water temperature was raised to 28 °C to simulate moderate hypothermia [1]. The temperatures were recorded every 30 seconds until the water bath reached a steady temperature of 28 °C as indicated by the immersion circulator. The temperature was maintained at 28 °C

Fig. 1. Experimental set up using a sous vide immersion circulator, transesophageal echocardiography probe, esophageal temperature probe, and bladder temperature probe in a 6-L water bath.

for 15 minutes, and temperatures were recorded every 60 seconds. This process was repeated for a temperature of 32 °C to represent mild hypothermia and 36 °C to represent normothermia [1]. Each time a target temperature was reached, a brief resuscitative TEE protocol was simulated [7], in which an operator performed an omniplane to 130°, saved an image, performed an omniplane back to 0°, and saved another image. This was all conducted without removing the probe from the water bath. Apart from this brief protocol, continuous 2D video monitoring without Doppler was maintained throughout the experiment.

Statistical analysis

Statistical analysis was performed in Microsoft Excel (Microsoft Corp). Bland-Altman plots were created to analyze the agreement between the sous vide immersion circulator temperature/water bath temperature and the TEE probe, bladder temperature probe, and esophageal temperature probe.

RESULTS

Continuous temperature monitoring was performed for a total of 70 minutes, with 78 temperature recordings for each device (Fig. 2). There were three episodes of water bath heating, each between 2 and 5 minutes. Fig. 3 displays the Bland-Altman plots of all three devices compared to the reported temperature of the sous vide immersion circulator. The difference between the immersion circulator and each device is plotted as a function of the mean of the two measurements. The TEE probe reported temperatures with a mean difference of 0.60 °C (95% confidence interval [CI], 0.51 to 0.69 °C). The upper level of agreement was 1.38 °C (95% CI, 1.23 to 1.54 °C) and the lower level of agreement was −0.18 °C (95% CI, −0.02 to −0.34°C) (Fig. 3A). Compared to the immersion circulator, the esophageal probe showed

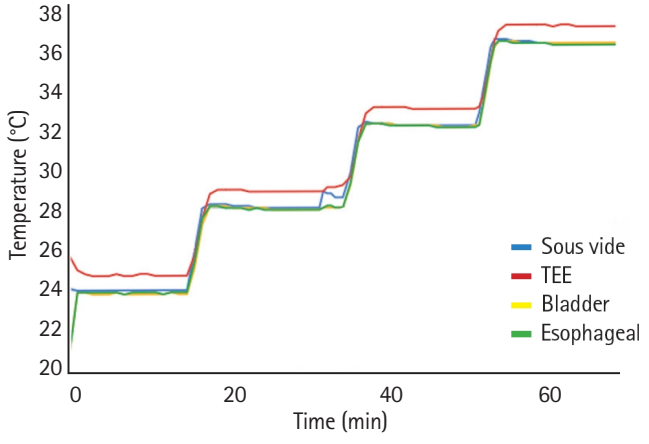


Fig. 2. Reported temperatures in a sous vide cooker-heated water bath over time by three medical-grade devices: a transesophageal echocardiography (TEE) transducer, a temperature-sensing bladder catheter, and a temperature-sensing esophageal probe.

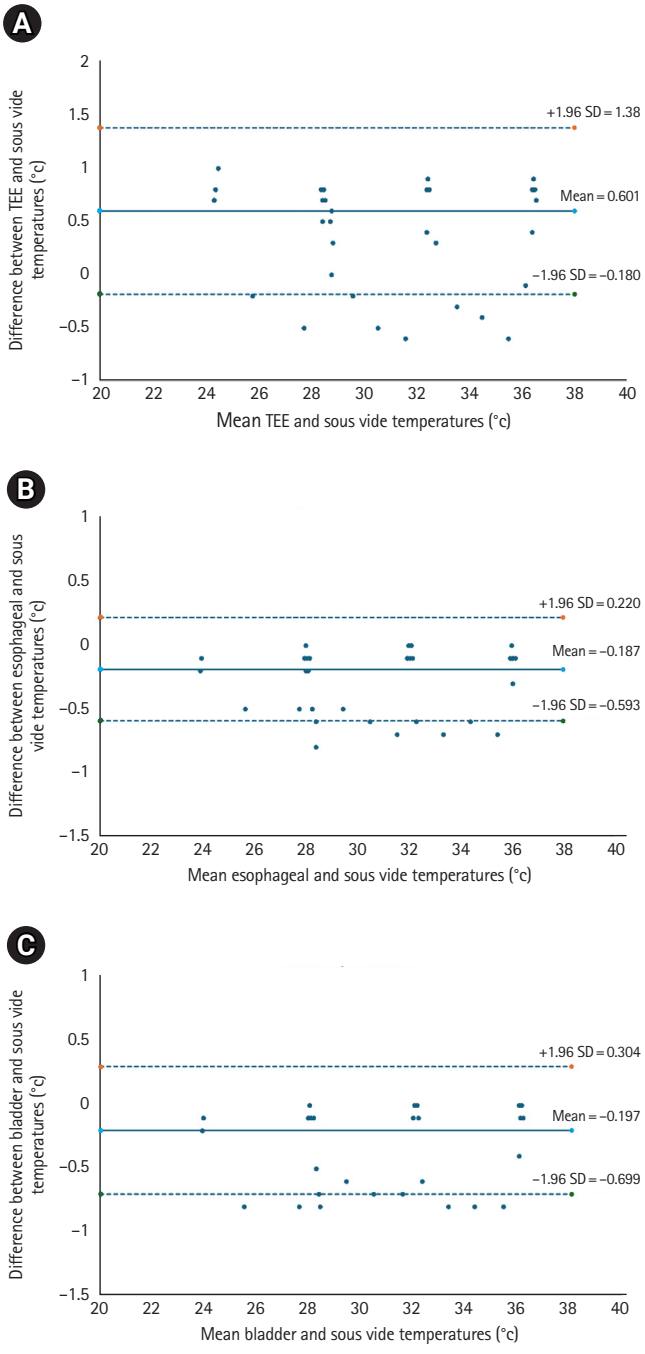


Fig. 3. Bland-Altman plots comparing the mean recorded temperature and the difference in recorded temperature between the sous vide immersion circulator and each medical device: (A) transesophageal echocardiography (TEE) probe, (B) esophageal temperature probe, and (C) bladder temperature probe. SD, standard deviation.

a mean difference of –0.19 °C (95% CI, –0.23 to –0.14 °C). The upper level of agreement was 0.22 °C (95% CI, 0.14 to 0.30 °C) and the lower level of agreement was –0.59 °C (95% CI, –0.67 to –0.51 °C) (Fig. 3B). For the bladder temperature probe, the mean temperature difference was –0.20 °C (95% CI, –0.26 to –0.14 °C). The upper level of agreement was 0.30 °C (95% CI, 0.20 to 0.41 °C) and the lower level of agreement was –0.70 °C (95% CI, –0.80 to –0.60 °C) (Fig. 3C). With the circulator as the reference standard, the TEE probe and both monitoring devices effectively tracked temperature changes. The mean TEE temperature was higher than the reference standard, while the mean esophageal and bladder probes were lower. The mean difference was lowest with the esophageal probe and highest with the TEE probe, but all modalities tracked temperature changes accurately.

DISCUSSION

Based on manufacturer testing, this TEE probe has the potential to heat accurately under certain circumstances [5]. During this simulation, using only 2D mode for an extended period of time and occasional omniplane, the TEE tip temperature did not increase unexpectedly other than changes from the water bath temperature. The probe temperature was an average of 0.6 °C warmer than the reported sous vide immersion circulator temperature and was less accurate than both the esophageal and bladder temperature probes. This temperature difference is unique to the TEE probe that was used in this experiment and may vary between TEE probe types. The thermometer on the TEE probe used in this experiment cannot be calibrated; if calibration was possible, temperature reporting might have been more accurate.

The mean difference in temperature between the TEE probe and the sous vide immersion circulator during rewarming episodes fell below the lower level of agreement of –0.18 °C, whereas the mean temperature difference during steady state water bath temperature was 0.6 °C, suggesting a TEE probe lag during rewarming phases. The TEE probe quickly reached a steady state temperature upon water bath equilibrium. Rewarming during hypothermic cardiac arrest typically is performed very slowly, so the lag in rapid temperature change is unlikely to be clinically relevant. The esophageal and bladder temperature probes also recorded temperatures lower than the reported water bath temperature during episodes of warming. This suggests that the water bath temperature was inaccurate or that these temperature probes also lagged.

This study has several limitations. The major limitation is that this is an experimental model using a water bath and might not be applicable to humans. This simulation tested a single TEE probe and was less reliable than simultaneous testing of multiple probes of the same design. Our findings cannot be generalized to other TEE probes on the market. There was no gold standard for the water bath temperature. While we assigned the sous vide immersion circulator temperature as the standard for comparison, this temperature may have been inaccurate, limiting the conclusions. Use of the Bland-Altman plot, even in the absence of a gold standard, offers insight into the relative accuracy and consistency of the different temperature probes. The water bath temperature also may have varied in different parts of the container, although efforts were made to position the devices in the same area of the water bath.

In this simulated model of hypothermia and rewarming, the TEE probe demonstrated precision in monitoring stable hypothermia and rewarming episodes but was less accurate than esophageal and bladder temperature probes. During an extended period of continuous 2D ultrasound without Doppler, the TEE probe did not heat up outside of the expected water bath temperature changes. Based on this study, TEE probes may not be ideal to determine an accurate initial core temperature but can likely be used to monitor changes in temperature over time. Future *in vivo* data may further inform the use of TEE to monitor rewarming.

ARTICLE INFORMATION

Author contributions

Conceptualization: ML, MP, DM, JR, CW; Data curation: ML, CW; Formal analysis: ML, CW; Investigation: ML, MP, DM, PC, CW; Methodology: ML, MP, DM, AF, PC, CW; Resources: ML, MP; Visualization: CW; Writing–original draft: ML, CW; Writing–review & editing: MP, DM, AF, PC, JR. All authors read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

Funding

The authors received no financial support for this study.

Acknowledgments

The authors thank Rachel Williams (Department of Emergency Medicine, MaineHealth Maine Medical Center Portland; Tufts University School of Medicine) for the use of her sous vide immersion circulator.

Data availability

Data analyzed in this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Paal P, Pasquier M, Darocha T, et al. Accidental hypothermia: 2021 update. Int J Environ Res Public Health 2022;19:501.
2. Hymczak H, Gołąb A, Mendrala K, et al. Core tempera ture measurement: principles of correct measurement, problems, and complications. Int J Environ Res Public Health 2021;18: 10606.
3. Dow J, Giesbrecht GG, Danzl DF, et al. Wilderness Medical Society clinical practice guidelines for the out-of-hospital evaluation and treatment of accidental hypothermia: 2019 update. Wilderness Environ Med 2019;30(4S): S47–69.
4. Teran F, West FM, Jelic T, et al. Resuscitative transesophageal echocardiography in emergency departments in the United States and Canada: a cross-sectional survey. Am J Emerg Med 2024;76:164–72.
5. Fujifilm SonoSite Inc. T8-3 Transducer (P29465) user guide. Fujifilm SonoSite Inc; 2023.
6. Misra S, Das PK, Srinivasan A. Performance of the transoesophageal echocardiography probe as an oesophageal temperature monitor in patients undergoing cardiac surgery with cardiopulmonary bypass: a prospective observational study. Eur J Cardiothorac Surg 2023;64:ezad242.
7. Teran F. Resuscitative cardiopulmonary ultrasound and transesophageal echocardiography in the emergency department. Emerg Med Clin North Am 2019;37:409–30.

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):86-89
https://doi.org/10.15441/ceem.24.217

Case Report

Mortality due to minimal bentazone intoxication: a case report

Yumin Jeon, Sejoong Ahn, Jong-Hak Park, Hanjin Cho, Sungwoo Moon, Sukyo Lee

Department of Emergency Medicine, Korea University Ansan Hospital, Ansan, Korea

pISSN: 3022-1811
eISSN: 2383-4625

Bentazone is a widely used herbicide considered a moderate hazard. Fatalities associated with its use are rare, with deaths reported only after intake of ≥200 mL. In some studies, bentazone intoxication is accompanied by generalized rigidity. Malignant hyperthermia is a pharmacogenetic disease that presents a hypermetabolic response to anesthetic gases or depolarizing muscle relaxants due to calcium channel dysfunction. The classic symptoms of malignant hyperthermia include hyperthermia and muscle rigidity. We report the case of a 65-year-old man who died 4 hours after presenting to the emergency department after ingesting approximately 75 mL of Basagran M60 (bentazone 33.6%, 25.2 g). This is the smallest dose (364 mg/kg) reported in a fatal case to date. Electrocardiogram changes, including QRS widening and QT prolongation, were present, and hypocalcemia was confirmed. We propose that bentazone intoxication causes patient deterioration by a mechanism similar to that of malignant hyperthermia.

Keywords Toxicology; Herbicides; Emergencies; Case reports

Received: March 13, 2024
Revised: June 13, 2024
Accepted: June 13, 2024

Correspondence to:
Sukyo Lee
(sukyolee@korea.ac.kr)

INTRODUCTION

Many cases of acute overdose in Korea are attributed to deliberate consumption of agricultural pesticides and herbicides as a means of suicide [1]. Bentazone is an acidic herbicide classified as moderately hazardous by the World Health Organization (WHO) [2]. Among the cases reported to date, only two have included fatal outcomes and muscle rigidity within a few hours [3,4]. We report a case of mortality after bentazone intoxication, with death occurring within 4 hours of ingestion.

Capsule Summary

What is already known

Bentazone is a widely used herbicide and is considered a moderate hazard. Fatalities are rarely reported, with reports of deaths occurring in doses of 200 mL or more.

What is new in the current study

To our knowledge, this is the case of bentazone intoxication resulting in mortality at the lowest reported dose. Also, this is the first to report that bentazone intoxication causes hypocalcemia, with electrocardiographic changes including QRS widening and QT prolongation.

How to cite this article:
Jeon Y, Ahn S, Park JH, Cho H, Moon S, Lee S. Mortality due to minimal bentazone intoxication: a case report. Clin Exp Emerg Med 2026;13(1):86-89. https://doi.org/10.15441/ceem.24.217

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

© 2026 The Korean Society of Emergency Medicine

CASE REPORT

A 65-year-old man visited the emergency department (ED) 40 minutes after ingesting approximately 75 mL of Basagran M60 (bentazone 33.6%, 25.2 g; FMC Corp) to commit suicide. The only co-ingestion was alcohol. He had no medical history.

The patient was alert and did not report any symptoms. His vital signs were as follows: blood pressure, 128/64 mmHg; heart rate, 106 beats/min; respiratory rate, 20 breaths/min; body temperature, 37.3 °C; and oxygen saturation, 94%. Results of his physical examination, including chest auscultation, were unremarkable.

Initial arterial blood gas analysis (ABGA) revealed respiratory alkalosis (pH, 7.482; PCO_2, 32.3 mmHg; PO_2, 68 mmHg; HCO_3, 24.2 mmol/L). Notably, his ionized calcium level was 1.4 mg/dL. Other abnormal laboratory results included lactic acid at 2.3 mmol/L, osmolality at 329 mOsmol/kg, and ammonia at 149 μg/dL. His serum potassium level was 3.9 mmol/L, and his creatine kinase concentration was 56 U/L. Additionally, his serum glucose concentration was 114 mg/dL, and his blood urea nitrogen and serum creatinine were 9 and 1.1 mg/dL, respectively. All other laboratory results were normal. Electrocardiography (ECG) revealed sinus tachycardia with significant QT prolongation (Fig. 1A). Findings on chest plain radiographs were normal. The patient was uncooperative, and it was not possible to perform gastric lavage or administer activated charcoal. Normal saline (500 mL) and sodium bicarbonate (60 mEq) were administered. Calcium replacement was not performed.

Approximately 90 minutes after ED admission, the patient had a body temperature of 38.7 °C. The patient was treated with evaporative external cooling and urinary bladder irrigation with cold saline. Acetaminophen (1,000 mg) was also intravenously injected. One hour later, his tachypnea was aggravated, with a respiratory rate exceeding 30 breaths/min and a rapid deterioration of mental status, with a Glasgow Coma Scale of 3 points. Subsequent ABGA revealed a pH of 7.555, PCO_2 of 24.2 mmHg, PO_2 of 81 mmHg, and HCO_3 of 21.0 mmol/L. Moreover, his lactate level had increased to 7.49 mmol/L. Brain computed tomography was performed because of his comatose state but showed no acute lesions. The patient's body temperature rose to 41.0 °C right after the computed tomography scan. An additional 1,000 mL of normal saline was administered.

After 163 minutes of ED admission, endotracheal intubation was performed for airway protection. The patient was intubated with 20 mg of etomidate without a neuromuscular blocker. At this time, he did not exhibit any muscle rigidity. His blood pressure subsequently decreased to 63/34 mmHg 5 minutes after endotracheal intubation. Despite the administration of norepinephrine up to 0.34 μg/kg/min, cardiac arrest occurred 10 minutes after endotracheal intubation, with the initial rhythm being pulseless electrical activity. ABGA before the arrest showed a pH of 7.237, PCO_2 of 56.7 mmHg, PO_2 of 90 mmHg, and HCO_3 of 24.1 mmol/L. His lactate level was 10.21 mmol/L, and an ionized calcium test yielded a result of < 1.0 mg/dL. Also, his serum potassium level was 5.5 mmol/L, and his serum glucose concentration was 41 mg/dL. Finally, his blood urea nitrogen and serum creatinine levels were 12 and 1.3 mg/dL, respectively. Creatine kinase was not rechecked. ECG showed QRS widening and QT prolongation (Fig. 1B).

Two minutes after the cardiac arrest, asystole was observed in every cycle. Despite resuscitation efforts, the patient died. Muscle rigidity was observed throughout his body at the time of death (Fig. 2).

Ethics statement

This study was approved by the Institutional Review Board of Korea University Ansan Hospital (No. 2024AS0012). Informed consent for publication of the research details and clinical images was obtained from the legal guardian of the patient.

DISCUSSION

Bentazone, a herbicide of the thiadiazine group, interferes with photosynthesis by preventing carbon dioxide fixation [5]. It is widely used for the cultivation of rice, perilla, barley, and corn [6]. The WHO classifies bentazone as moderately hazardous, and fatal cases following exposure are very rare [2]. Muller et al. [7] previously suggested that a dose of 686 to 1,371 mg/kg is lethal in humans. Fatal cases of liver failure, kidney injury, and muscle rigidity with failed or delayed airway access have been reported [3–7]. Of the five fatal cases of acute bentazone intoxication reported to date, the smallest dose was 100 to 200 mL (48%), while all other cases involved doses >200 mL [3–7]. In the presented case, the dose was 364 mg/kg.

Previous studies reported delayed airway establishment due to muscle rigidity. Lee [3] and Turcant et al. [4] reported no improvement in rigidity despite succinylcholine administration, suggesting a need for early surgical airway. However, in this case, despite achieving a definite airway before muscle rigidity developed, the patient eventually expired. This suggests that the cause of death is something other than respiratory failure, and the recommendation of early advanced airway is questionable.

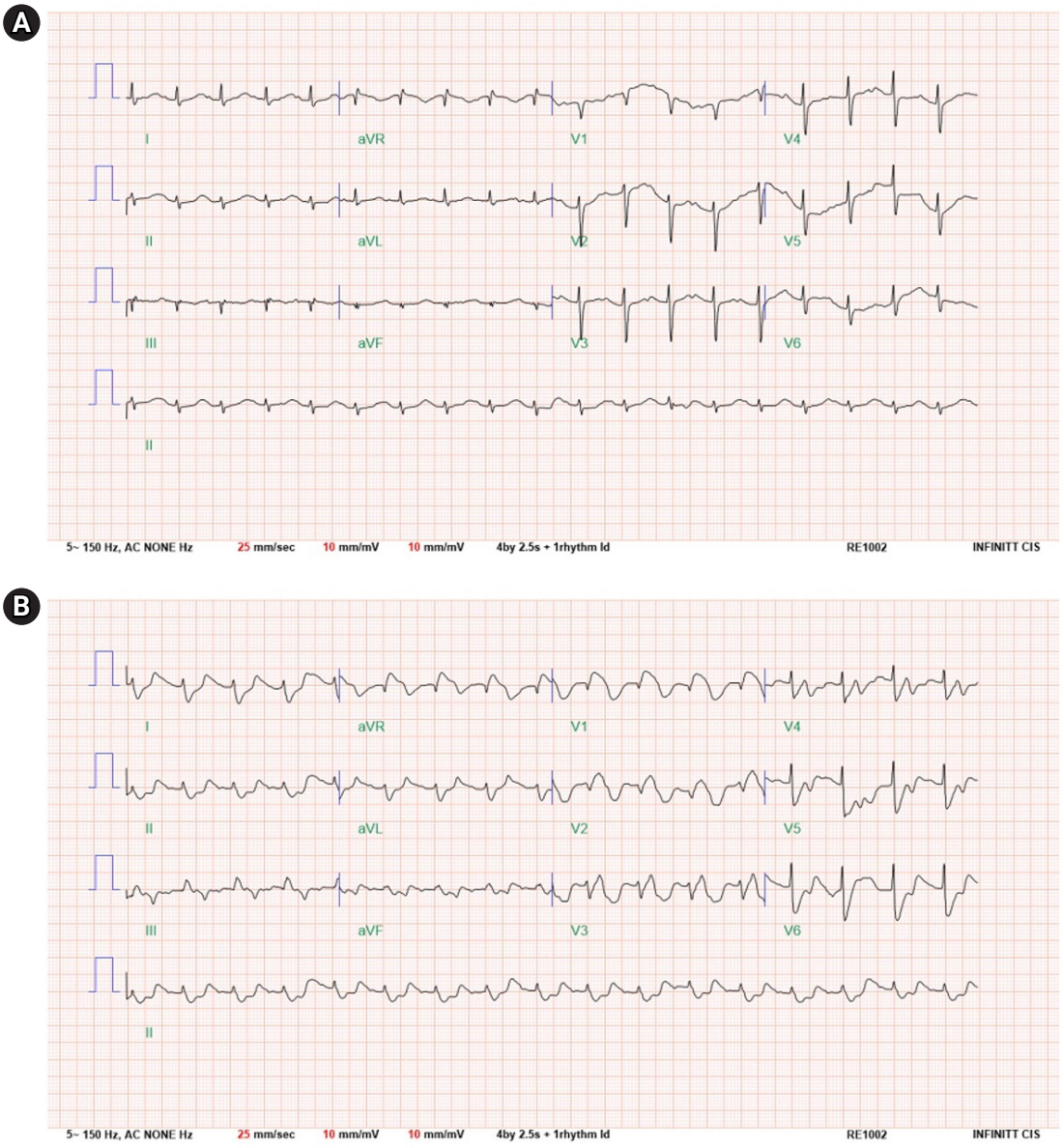


Fig. 1. Electrocardiogram of the patient. (A) One hundred one minutes after emergency department admission. The patient's corrected QT was 615 msec. (B) Two hundred thirty-five minutes after emergency department admission. The patient's QRS duration was 184 msec.

Neuroleptic malignant syndrome (NMS) and malignant hyperthermia (MH) are similar in that they cause symptoms such as fever, muscle rigidity, and autonomic dysfunction. Lin et al. [8] reported that bentazone intoxication mimics NMS. However, we propose that bentazone intoxication mimics MH rather than NMS. First, MH occurs in a small percentage of genetically susceptible individuals. While bentazone is classified as moderately hazardous, and most patients have an insidious course after consumption, a few patients have been reported to develop muscle rigidity and a fatal course. The reported fatal cases may have been genetically susceptible to bentazone. Second, hypocalcemia and QT prolongation in our patient were consistent with the pathophysiology of MH, which is based on calcium release from the sarcoplasmic reticulum and extracellular calcium influx. This indicates that bentazone may induce a potential influx of extracellular calcium. In this regard, patients with bentazone intoxication and an anticipated fatal course may be candidates for dantrolene. Dantrolene inhibits calcium ions release from the sarcoplasmic reticulum by antagonizing ryanodine receptors. Therefore, it is a key part of the treatment of MH and is also used in other conditions that cause muscle contraction and heat production, including NMS. Despite hypocalcemia, the benefit of calcium ad-

Fig. 2. Clinical photograph after termination of resuscitation. The patient developed generalized rigidity due to bentazone intoxication at the time of death. His left arm is rigid, with the elbow flexed and the hand in the air, and both of his feet are rigid, with plantar flexion. The patient's legal guardian provided informed consent for publication of the clinical image.

ministration is questionable because it may increase the intracellular calcium concentration, which may lead to a fatal outcome. Also, since succinylcholine is one of the causes of MH, it should be avoided.

To our knowledge, this case report documents the lowest dose of bentazone that led to a fatal outcome to date. It is also the first to report hypocalcemia and ECG changes following bentazone intoxication. In addition, we suggest an association between bentazone intoxication and MH. A limitation of our case is that, although we propose dantrolene as a therapeutic candidate, it has not been applied in practice. Future studies are required to validate this therapeutic effect. Furthermore, it will be necessary to investigate the presence of ryanodine receptor mutations to confirm that those who have been fatally affected by bentazone intoxication indeed have MH.

ARTICLE INFORMATION

Author contributions

Conceptualization: all authors; Formal analysis: all authors; Investigation: YJ; Methodology: all authors; Supervision: SL; Writing–original draft: YJ; Writing–review & editing: SA, JHP, HC, SM, SL. All authors read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

Funding

The authors received no financial support for this study.

Data availability

Data sharing is not applicable as no new data were created or analyzed in this study.

REFERENCES

1. Sung AJ, Lee KW, So BH, et al. Multicenter survey of intoxication cases in Korean emergency departments: 2nd annual report, 2009. J Korean Soc Clin Toxicol 2012;10:22–32.
2. World Health Organization (WHO). The WHO recommended classification of pesticides by hazard and guidelines to classification: 2009. WHO; 2010.
3. Lee K. A fatal case of acute bentazone overdose despite cricothyroidotomy during cardiopulmonary resuscitation. Clin Exp Emerg Med 2017;4:254–7.
4. Turcant A, Harry P, Cailleux A, et al. Fatal acute poisoning by bentazon. J Anal Toxicol 2003;27:113–7.
5. Wu IW, Wu MS, Lin JL. Acute renal failure induced by bentazone: 2 case reports and a comprehensive review. J Nephrol 2008;21:256–60.
6. Skavic P, Durakovic Z, Nestic M. Lethal bentazone intoxication: a report of two cases. J Forensic Sci 2018;63:321–5.
7. Muller IB, Willads Petersen H, Johansen SS, Theilade P. Fatal overdose of the herbicide bentazone. Forensic Sci Int 2003;135:235–6.
8. Lin TJ, Hung DZ, Hu WH, Yang DY, Wu TC. Acute basagran poisoning mimicking neuroleptic malignant syndrome. Hum Exp Toxicol 1999;18:493–4.

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):90-97
https://doi.org/10.15441/ceem.25.318

Commentary

pISSN: 3022-1811
eISSN: 2383-4625

Epidemiology of cardiac arrest and acute myocardial infarction in emergency departments in Korea, 2020–2024: a report using the National Emergency Department Information System (NEDIS) database

Sejoong Ahn[1*], Jung Eon Kim[2*], Sung-keun Ko[2], Eunsil Ko[2]

[1]Department of Emergency Medicine, Korea University Ansan Hospital, Ansan, Korea
[2]National Emergency Medical Center, National Medical Center, Seoul, Korea

Received: December 18, 2025
Revised: January 6, 2026
Accepted: January 6, 2026

Correspondence to:
Sejoong Ahn
(sejoongahn@naver.com)

Co-correspondence to:
Jung Eon Kim
(jekim1229@nmc.or.kr)

*Sejoong Ahn and Jung Eon Kim contributed equally to this study as co-first authors.

INTRODUCTION

Cardiac arrest and acute myocardial infarction are major time-sensitive and resource-intensive emergencies encountered in emergency departments (EDs). Both conditions require rapid diagnosis, high-quality resuscitation, coordinated multidisciplinary responses, and timely access to advanced interventions, all of which are strongly associated with improved patient outcomes [1–3]. In addition, acute myocardial infarction is a major underlying cause of cardiac arrest and necessitates prompt reperfusion therapy, including coronary angiography and percutaneous coronary intervention, often within narrow therapeutic windows [1–3]. Therefore, the epidemiologic characteristics of these two interrelated conditions in the ED not only reflect the burden of acute cardiovascular emergencies but also serve as important indicators of the performance and resilience of the emergency care system.

This study summarizes the nationwide epidemiology of cardiac arrest and acute myocardial infarction in EDs from 2020 to 2024. It provides updated insights into their incidence, outcomes, and temporal patterns across age groups. Understanding these trends in the post–COVID-19 period, as well as during the 2024 healthcare policy dispute in Korea [4], is essential for optimizing resource allocation, strengthening emergency preparedness, and informing health policies aimed at improving outcomes for critically ill patients.

METHODS

Ethics statement

This study was approved by the Institutional Review Board of the National Medical Center, Korea (No. NMC-2023-08-094). The requirement for informed consent was waived due to the retrospective, observational, and anonymous nature of the study.

Study population

This study used nationwide data from Korea's National Emergency Department Information Sys-

How to cite this article:
Ahn S, Kim JE, Ko S, Ko E. Epidemiology of cardiac arrest and acute myocardial infarction in emergency departments in Korea, 2020–2024: a report using the National Emergency Department Information System (NEDIS) database. Clin Exp Emerg Med 2026;13(1):90-97. https://doi.org/10.15441/ceem.25.318

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

© 2026 The Korean Society of Emergency Medicine

tem (NEDIS). NEDIS is a government-operated registry that collects standardized administrative and clinical information from all designated EDs in Korea [5]. Cardiac arrest was defined as patients identified by cardiopulmonary resuscitation–related codes in ED management records or ED discharge diagnoses, excluding patients who were dead on arrival (DOA). This definition included both out-of-hospital and in-hospital cardiac arrest cases and encompassed all ED levels (level I, regional emergency medical center; level II, local emergency medical center; level III, local emergency medical institution). Acute myocardial infarction was defined as patients presenting to the ED within 24 hours of symptom onset who had an ED or hospital discharge diagnosis code of I21, excluding DOA cases, and was limited to level I and level II EDs. Patient demographics, prehospital characteristics, and hospital-level characteristics were extracted for analysis.

Statistical analysis

Descriptive statistics were calculated. Age- and sex-standardized ED visit rates per 100,000 population were calculated using the 2020 mid-year estimated population of Korea as the standard population. Age- and sex-standardized mortality rates per 100,000 ED visits were calculated using the number of ED visits recorded in NEDIS from 2020 to 2024 as the standard population. Linear regression analysis was performed to evaluate 5-year trends. SAS ver. 9.4 (SAS Institute Inc) was used for all statistical analyses.

EPIDEMIOLOGY OF CARDIAC ARREST IN EDs, 2020–2024

Table 1 presents the characteristics of patients with cardiac arrest in the ED. From 2020 to 2024, a total of 205,617 patients with cardiac arrest visited EDs nationwide (pediatric group [0–17 years], 1.6%; adult group [18–64 years], 36.4%; older adult group [≥65 years], 62.0%). The mean age was 67.4 years, and the median time from symptom onset to ED arrival was 40 minutes. Pediatric patients with cardiac arrest visited level I EDs more frequently, whereas older adults more often visited level III EDs compared with other age groups. Most patients arrived at the ED as direct visits (87.3%) and were transported by 119 ambulances (80.9%), as expected. The mean and median ED length of stay (LOS) were 4.0 and 1.9 hours, respectively, with the shortest mean LOS observed in the pediatric group. The overall in-hospital mortality rate was 77.6%, with the highest mortality observed among older adults (pediatric group, 71.8%; adult group, 69.2%; older adult group, 82.6%).

Age- and sex-standardized ED visit rates ranged from 69.8 to 78.3 per 100,000 population, with no significant changes observed over the study period (adults, 40.8–43.3 per 100,000 population; older adults, 254.5–298.3 per 100,000 population) (Fig. 1A). Notably, age- and sex-standardized ED visit rates for pediatric cardiac arrest increased significantly from 7.6 per 100,000 population in 2020 to 10.7 per 100,000 population in 2024 (P for trend = 0.028). Age- and sex-standardized mortality rates per 100,000 ED visits showed no significant temporal changes (overall, 378.9–467.5; adult group, 221.6–265.9; pediatric group, 33.1–50.3; older adult group, 437.4–553.3) (Fig. 1B). Similarly, in-hospital mortality rates remained stable over time (overall, 76.4%–78.7%; adult group, 68.6%–70.8%; pediatric group, 68.1%–74.0%; older adult group, 81.5%–83.5%) (Fig. 1C).

Age- and sex-standardized ED visit rates modestly declined from 2023 to 2024, decreasing from 73.9 to 69.8 per 100,000 population. In contrast, age- and sex-standardized mortality per 100,000 ED visits increased across all age groups, with the overall rate rising from 378.9 in 2023 to 430.2 in 2024. In-hospital mortality rates showed no significant year-to-year change across any subgroup, remaining stable overall (76.4% in 2023 to 76.9% in 2024).

EPIDEMIOLOGY OF ACUTE MYOCARDIAL INFARCTION IN EDs, 2020–2024

Table 2 shows the characteristics of patients with acute myocardial infarction in the ED. A total of 142,499 patients with acute myocardial infarction visited the ED during the study period (pediatric group, 0.1%; adult group, 45.0%; older adult group, 54.9%), with a mean age of 66.5 years. The median time from symptom onset to ED arrival was 133 minutes. Most patients visited the ED as direct visits (70.5%), while 27.0% were transferred from other hospitals. Regarding transport methods, 46.9% of patients arrived at the ED via 119 ambulances, and 34.4% arrived via other transport methods. The mean and median ED LOS were 5.9 and 3.2 hours, respectively. The overall in-hospital mortality rate was 9.6%.

When patient characteristics were assessed by age group, the median time from symptom onset to ED arrival was shortest in the adult group. Pediatric patients more frequently visited level II EDs, were more likely to present as direct visits, and less frequently used ambulance transport compared with other age groups. The in-hospital mortality rate was highest among older adults (pediatric group, 1.3%; adult group, 4.5%; older adult group, 13.7%).

Table 1. Demographics, ED visit characteristics, and outcomes of patients with cardiac arrest

Variable	Total (n = 205,617)	Pediatric group (0–17 yr) (n = 3,325, 1.6%)	Adult group (18–64 yr) (n = 74,751, 36.4%)	Older adult group (≥ 65 yr) (n = 127,535, 62.0%)
Age (yr)	67.4 ± 18.0	7.5 ± 6.5	50.8 ± 11.3	78.7 ± 7.8
Sex				
Male	130,953 (63.7)	1,935 (58.2)	54,994 (73.6)	74,019 (58.0)
Female	74,664 (36.3)	1,390 (41.8)	19,757 (26.4)	53,516 (42.0)
Time from symptom onset to ED arrival[a]	40 (28–74)	36 (25–66)	40 (28–73)	40 (29–74)
Type of ED				
Level I	67,305 (32.7)	1,606 (48.3)	26,265 (35.1)	39,434 (30.9)
Level II	98,648 (48.0)	1,408 (42.3)	36,107 (48.3)	61,133 (47.9)
Level III	39,664 (19.3)	311 (9.4)	12,379 (16.6)	26,968 (21.1)
Route of arrival				
Direct visit	179,417 (87.3)	2,924 (87.9)	65,008 (87.0)	111,479 (87.4)
Transfer from other hospital	22,808 (11.1)	362 (10.9)	8,334 (11.1)	14,112 (11.1)
Referral from outpatient clinic	3,178 (1.5)	39 (1.2)	1,336 (1.8)	1,803 (1.4)
Other	112 (0.1)	0 (0)	37 (0.0)	75 (0.1)
Unknown	102 (0.0)	0 (0)	36 (0.0)	66 (0.1)
Transport				
119 Ambulance	166,401 (80.9)	2,646 (79.6)	59,960 (80.2)	103,790 (81.4)
Other medical institution ambulance	4,551 (2.2)	73 (2.2)	1,577 (2.1)	2,901 (2.3)
Other ambulance	14,608 (7.1)	165 (5.0)	5,214 (7.0)	9,229 (7.2)
Police or official transport	225 (0.1)	4 (0.1)	159 (0.2)	62 (0.0)
Air transport	881 (0.4)	16 (0.5)	478 (0.6)	387 (0.3)
Other transport	17,674 (8.6)	402 (12.1)	6,967 (9.3)	10,304 (8.1)
Walk-in	441 (0.2)	11 (0.3)	163 (0.2)	267 (0.2)
Other/unknown	836 (0.4)	8 (0.2)	233 (0.3)	595 (0.5)
Length of stay (hr)[b]				
Mean ± SD	4.0 ± 7.6	3.0 ± 4.5	4.2 ± 7.7	4.0 ± 7.5
Median (IQR)	1.9 (1.0–3.9)	1.8 (1.1–3.1)	2.0 (1.1–4.1)	1.8 (1.0–3.8)
0–6	174,042 (84.6)	3,009 (90.5)	62,766 (84.0)	108,262 (84.9)
6–12	17,947 (8.7)	217 (6.5)	6,777 (9.1)	10,953 (8.6)
12–24	9,003 (4.4)	74 (2.2)	3,490 (4.7)	5,439 (4.3)
≥ 24	4,453 (2.2)	25 (0.8)	1,666 (2.2)	2,762 (2.2)
Unknown	172 (0.1)	0 (0)	52 (0.1)	119 (0.1)
ED disposition				
Discharge	6,016 (2.9)	106 (3.2)	3,236 (4.3)	2,673 (2.1)
Admission[c]	54,620 (26.6)	903 (27.2)	24,517 (32.8)	29,200 (22.9)
General ward	5,130 (2.5)	103 (3.1)	2,100 (2.8)	2,927 (2.3)
Intensive care unit	49,451 (24.0)	800 (24.1)	22,405 (30.0)	26,246 (20.6)
Transfer	11,598 (5.6)	335 (10.1)	5,253 (7.0)	6,009 (4.7)
Comfort care discharge	114 (0.1)	0 (0)	38 (0.1)	76 (0.1)
Death	133,077 (64.7)	1,975 (59.4)	41,638 (55.7)	89,460 (70.1)
Other/unknown	192 (0.1)	6 (0.2)	69 (0.1)	117 (0.1)
Comfort care disposition				
Discharge	24,830 (12.1)	489 (14.7)	13,660 (18.3)	10,680 (8.4)
Transfer	19,197 (9.3)	418 (12.6)	8,603 (11.5)	10,175 (8.0)
Comfort care discharge	321 (0.2)	6 (0.2)	122 (0.2)	193 (0.2)
Death	159,527 (77.6)	2,388 (71.8)	51,765 (69.2)	105,370 (82.6)
Other/unknown	1,742 (0.8)	24 (0.7)	601 (0.8)	1,117 (0.9)

Values are presented as mean±SD, number (%), or median (IQR). Differences between totals and subgroup sums are due to missing age data (n=6); percentages may not total 100 due to rounding.

ED, emergency department; SD, standard deviation; IQR, interquartile range.

[a]Reporting was not obligatory for level III EDs. [b]Any ED length of stay over 5 days was treated as 5 days. [c]Admission locations other than general wards or intensive care units were unspecified.

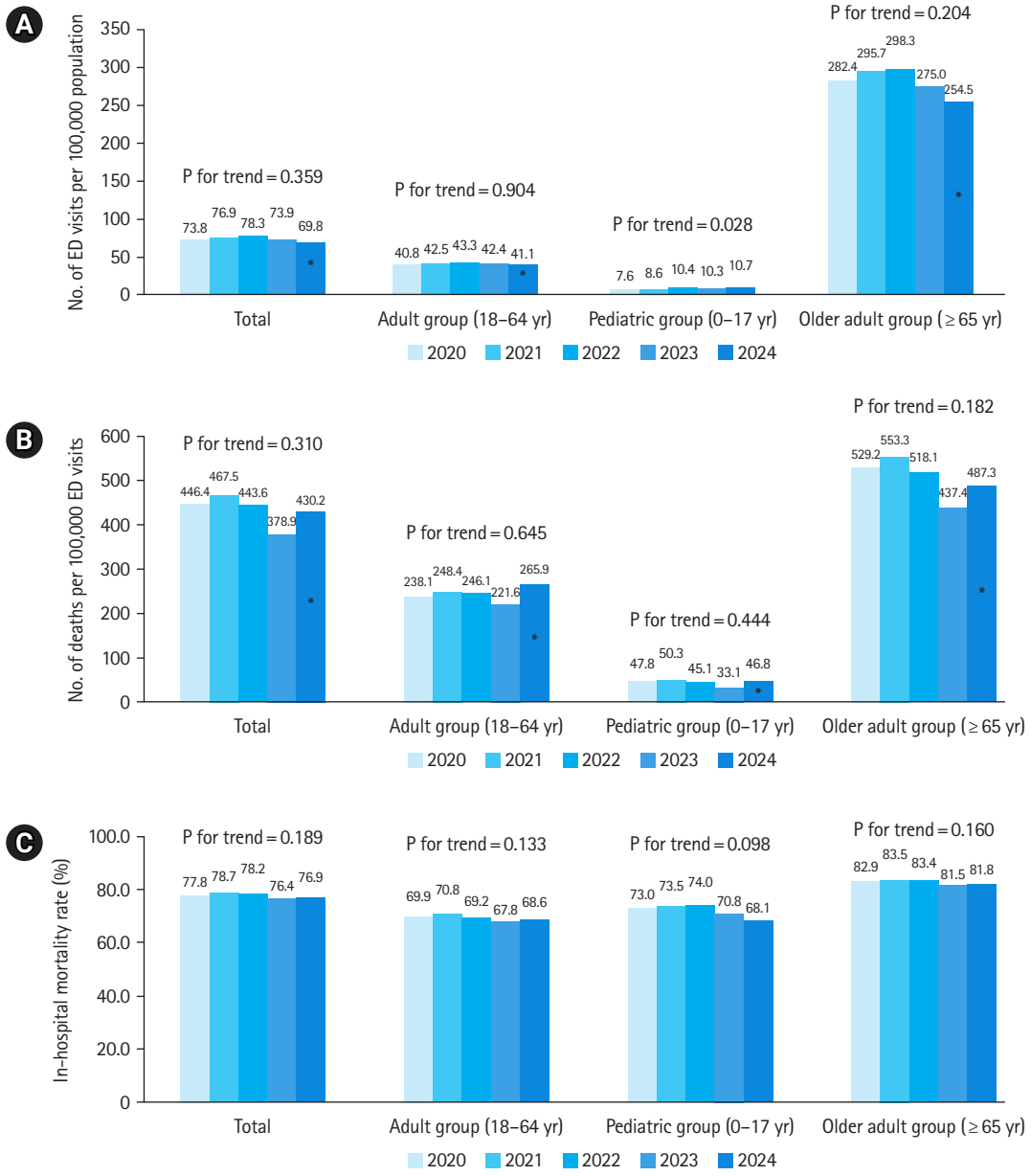


Fig. 1. Changes in the incidence and mortality of patients with cardiac arrest during 2020–2024. (A) Age- and sex-standardized emergency department (ED) visits per 100,000 population (2023 vs. 2024: total, 73.9 vs. 69.8, P<0.001; adult group, 42.4 vs. 41.1, P=0.006; pediatric group, 10.3 vs. 10.7, P=0.437; older adult group, 275.0 vs. 254.5, P<0.001). (B) Age- and sex-standardized mortality per 100,000 ED visits (2023 vs. 2024: total, 378.9 vs. 430.2, P<0.001; adult group, 221.6 vs. 265.9, P<0.001; pediatric group, 33.1 vs. 46.8, P<0.001; older adult group, 437.4 vs. 487.3, P<0.001). (C) In-hospital mortality (2023 vs. 2024: total, 76.4% vs. 76.9%, P=0.128; adult group, 67.8% vs. 68.6%, P=0.135; pediatric group, 70.8% vs. 68.1%, P=0.289; older adult group, 81.5% vs. 81.8%, P=0.427). *Statistically significant differences between 2023 and 2024.

Age- and sex-standardized ED visit rates ranged from 47.2 to 53.6 per 100,000 population, with no significant changes observed over the study period (adult group, 32.8–35.6 per 100,000 population; pediatric group, 0.1–0.4 per 100,000 population; older adult group, 153.7–181.4 per 100,000 population) (Fig. 2A). Overall age- and sex-standardized mortality per 100,000 ED visits significantly decreased over time (33.5–41.6 per 100,000 ED visits, P for trend = 0.018) (Fig. 2B), with the most pronounced decrease observed among older adults (38.8 to 34.2 per 100,000 ED visits, P for trend = 0.012). In contrast, age- and sex-standardized mortality per 100,000 ED visits among pediatric and adult patients showed no significant temporal changes (adult group, 12.7–14.6 per 100,000 ED visits; pediatric group, 0.0–0.1 per 100,000 ED visits). Similarly, overall in-hospital mortality signifi-

Table 2. Demographics, ED visit characteristics, and outcomes of patients with acute myocardial infarction

Variable	Total (n = 142,499)	Pediatric group (0–17 yr) (n = 75, 0.1%)	Adult group (18–64 yr) (n = 64,175, 45.0%)	Older adult group (≥ 65 yr) (n = 78,249, 54.9%)
Age (yr)	66.5 ± 13.7	11.6 ± 4.6	54.1 ± 7.9	76.7 ± 7.7
Sex				
Male	104,228 (73.1)	56 (74.7)	56,145 (87.5)	48,027 (61.4)
Female	38,271 (26.9)	19 (25.3)	8,030 (12.5)	30,222 (38.6)
Time from symptom onset to ED arrival	133 (60–360)	146.0 (60–420)	111.0 (53–297)	161 (60–420)
Type of ED				
Level I	62,684 (44.0)	16 (21.3)	27,609 (43.0)	35,059 (44.8)
Level II	79,815 (56.0)	59 (78.7)	36,566 (57.0)	43,190 (55.2)
Route of arrival				
Direct visit	100,407 (70.5)	67 (89.3)	46,528 (72.5)	53,812 (68.8)
Transfer from other hospital	38,429 (27.0)	7 (9.3)	16,128 (25.1)	22,294 (28.5)
Referral from outpatient clinic	3,643 (2.6)	1 (1.3)	1,514 (2.4)	2,128 (2.7)
Other	20 (0.0)	0 (0)	5 (0.0)	15 (0.0)
Unknown	0 (0)	0 (0)	0 (0)	0 (0)
Transport				
119 Ambulance	66,846 (46.9)	13 (17.3)	28,742 (44.8)	38,091 (48.7)
Other medical institution ambulance	5,672 (4.0)	2 (2.7)	2,195 (3.4)	3,475 (4.4)
Other ambulance	18,971 (13.3)	4 (5.3)	7,336 (11.4)	11,631 (14.9)
Police or official transport	104 (0.1)	0 (0)	75 (0.1)	29 (0.0)
Air transport	627 (0.4)	0 (0)	288 (0.4)	339 (0.4)
Other transport	49,066 (34.4)	56 (74.7)	24,969 (38.9)	24,041 (30.7)
Walk-in	983 (0.7)	0 (0)	487 (0.8)	496 (0.6)
Other/unknown	230 (0.2)	0 (0)	83 (0.1)	147 (0.2)
Length of stay (hr)[a]				
Mean ± SD	5.9 ± 8.0	3.7 ± 2.8	5.1 ± 7.4	6.5 ± 8.4
Median (IQR)	3.2 (1.4–6.7)	3.0 (2.1–4.5)	2.6 (1.1–5.7)	3.7 (1.8–7.4)
0–6	102,596 (72.0)	67 (89.3)	49,010 (76.4)	53,519 (68.4)
6–12	21,802 (15.3)	6 (8.0)	8,293 (12.9)	13,503 (17.3)
12–24	12,856 (9.0)	2 (2.7)	4,942 (7.7)	7,912 (10.1)
≥ 24	5,244 (3.7)	0 (0)	1,929 (3.0)	3,315 (4.2)
Unknown	1 (0.0)	0 (0)	1 (0.0)	0 (0)
ED disposition				
Discharge	5,241 (3.7)	29 (38.7)	2,419 (3.8)	2,793 (3.6)
Admission[b]	128,307 (90.0)	34 (45.3)	58,425 (91.0)	69,848 (89.3)
General ward	38,354 (26.9)	24 (32.0)	15,964 (24.9)	22,366 (28.6)
Intensive care unit	89,886 (63.1)	10 (13.3)	42,443 (66.1)	47,433 (60.6)
Transfer	6,299 (4.4)	11 (14.7)	2,684 (4.2)	3,604 (4.6)
Comfort care discharge	17 (0.0)	0 (0)	2 (0.0)	15 (0.0)
Death	2,582 (1.8)	1 (1.3)	619 (1.0)	1,962 (2.5)
Other/unknown	53 (0.0)	0 (0)	26 (0.0)	27 (0.0)
Comfort care disposition				
Discharge	114,529 (80.4)	63 (84.0)	56,493 (88.0)	57,973 (74.1)
Transfer	13,430 (9.4)	11 (14.7)	4,443 (6.9)	8,976 (11.5)
Comfort care discharge	117 (0.1)	0 (0)	26 (0.0)	91 (0.1)
Death	13,650 (9.6)	1 (1.3)	2,915 (4.5)	10,734 (13.7)
Other/unknown	773 (0.5)	0 (0)	298 (0.5)	475 (0.6)

Values are presented as mean ± SD, number (%), or median (IQR). Percentages may not total 100 due to rounding.

ED, emergency department; SD, standard deviation; IQR, interquartile range.

[a]Any ED length of stay over 5 days was treated as 5 days. [b]Admission locations other than general wards or intensive care units were unspecified.

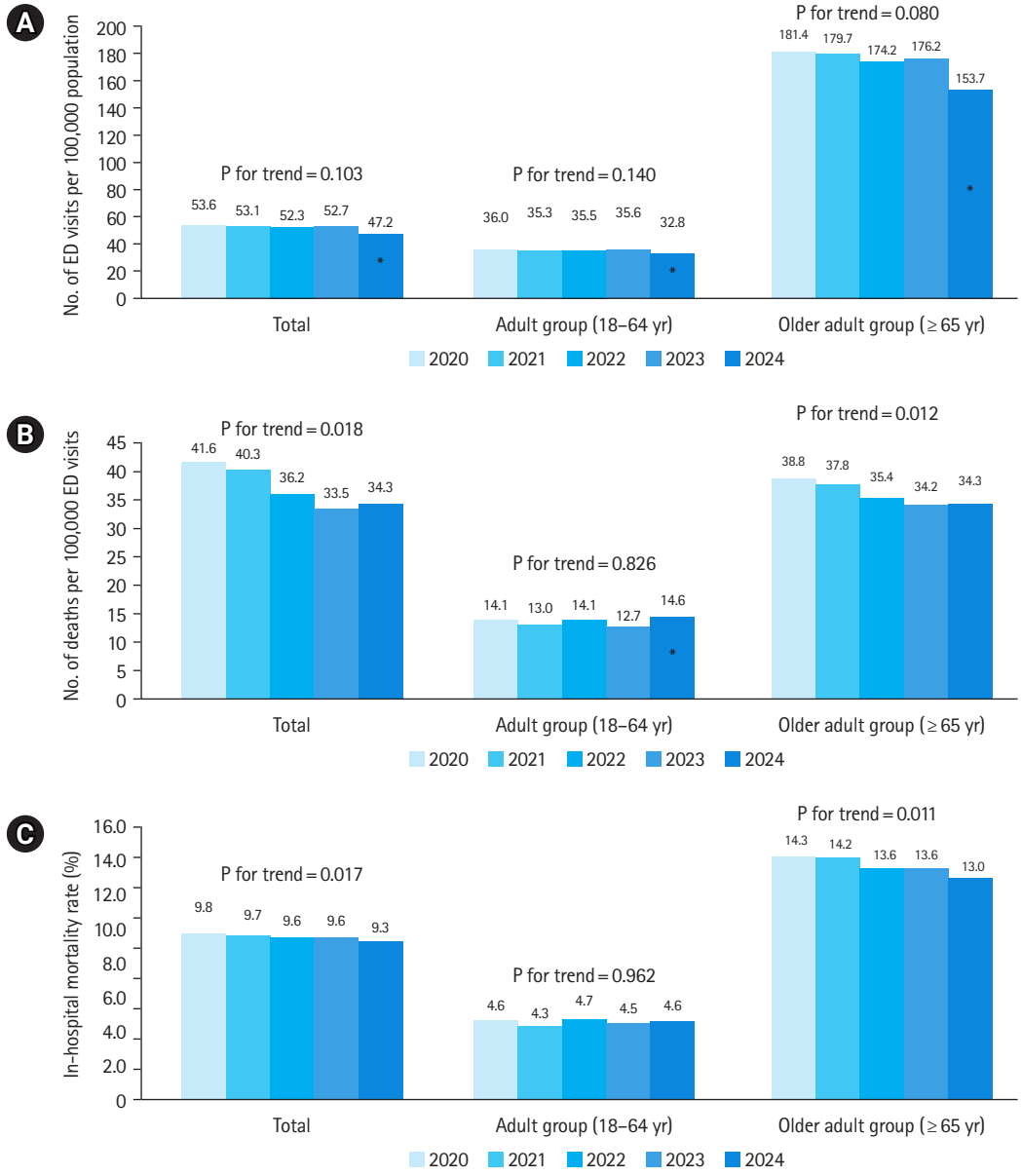


Fig. 2. Changes in the incidence and mortality of patients with acute myocardial infarction during 2020–2024. Pediatric patients with acute myocardial infarction are not presented in the figure because of the extremely small number of cases. (A) Age- and sex-standardized emergency department (ED) visits per 100,000 population (2023 vs. 2024: total, 52.7 vs. 47.2, P<0.001; adult group, 35.6 vs. 32.8, P=0.006; pediatric group, 0.2 vs. 0.1, P=0.424; older adult group, 176.2 vs. 153.7, P<0.001). (B) Age- and sex-standardized mortality per 100,000 ED visits (2023 vs. 2024: total, 33.5 vs. 34.3, P=0.378; adult group, 12.7 vs. 14.6, P=0.018; pediatric group, 0.0 vs. 0.0, P=not determined; older adult group, 34.2 vs. 34.3, P=0.978). (C) In-hospital mortality (2023 vs. 2024: total, 9.6% vs. 9.3%, P=0.264; adult group, 4.5% vs. 4.6%, P=0.668; pediatric group, 0.0% vs. 0.0%, P=not determined; older adult group, 13.6% vs. 13.0%, P=0.129). *Statistically significant differences between 2023 and 2024.

cantly decreased over time (9.8% to 9.3%, P for trend = 0.017), particularly among older adults (14.3% to 13.0%, P for trend = 0.011), whereas no significant changes were observed in pediatric or adult patients (Fig. 2C).

From 2023 to 2024, age- and sex-standardized ED visit rates decreased from 52.7 to 47.2 per 100,000 population. Over the same period, age- and sex-standardized mortality per 100,000 ED visits remained largely stable overall (33.5 in 2023 to 34.3 in 2024). In-hospital mortality showed no significant year-to-year change, both overall and across all subgroups (9.6% in 2023 to 9.3% in 2024).

DISCUSSION

This study summarizes the recent epidemiology of cardiac arrest and acute myocardial infarction in EDs between 2020 and 2024. The incidence of both conditions remained largely unchanged during the study period. While mortality associated with cardiac arrest showed no significant change, mortality related to acute myocardial infarction demonstrated a gradual improvement over time.

The overall incidence of cardiac arrest observed in this study was comparable to that reported worldwide [6]. Compared with a previous analysis of NEDIS data from 2018 to 2022 [7], the total number of cardiac arrest patients increased across all age groups. This increase likely reflects both methodological and demographic factors. The previous study restricted inclusion to patients presenting within 24 hours of symptom onset, which resulted in missing data from level III EDs. In contrast, the broader inclusion criteria applied in this study captured a substantial number of patients treated in level III EDs, thereby providing a more comprehensive representation of real-world emergency care. In addition, Korea's rapidly aging population likely contributed to the increased number of cardiac arrest cases among older adults [8,9]. Given the higher mortality observed in older adults and the expected continued growth of this population, both the incidence and mortality of cardiac arrest may increase gradually, placing a substantial burden on the acute cardiovascular emergency system. Moreover, the observed increase in pediatric cardiac arrest cases, although small in absolute numbers, warrants attention because of its potential socioeconomic impact and broader implications.

The overall number of patients with acute myocardial infarction and their epidemiological characteristics were similar to those reported in a previous NEDIS-based study covering 2018–2022 [10]. The observed decrease in mortality may reflect either genuine improvement in clinical outcomes or normalization of outcomes following the COVID-19 pandemic. During the pandemic, delayed hospital arrival, ED overcrowding, shortages of medical resources, and increased workload among medical staff were prominent and were associated with increased mortality from ischemic heart disease and cardiac arrest [11–14]. The normalization of these factors likely contributed to the observed reduction in mortality. However, further studies are required to determine whether this trend represents a true long-term improvement in outcomes for acute myocardial infarction.

From 2023 to 2024, in-hospital mortality remained largely unchanged. Although age- and sex-standardized mortality per 100,000 ED visits increased slightly during this period, standardized ED visit rates declined concurrently. This increase in standardized mortality per 100,000 ED visits may plausibly be attributed to a reduction in the total number of ED visits, which could have resulted in fewer standardized ED visits among patients with cardiac arrest and acute myocardial infarction. Importantly, despite a substantial reduction in the number of physicians in 2024 [4], in-hospital mortality did not worsen, suggesting that resource allocation and emergency preparedness were effectively maintained, at least within acute cardiovascular emergency systems.

Several limitations of this study should be acknowledged. First, clinically important variables, including Utstein variables, detailed management information, and neurological outcomes, were unavailable in the dataset. Second, level III ED data were not included for acute myocardial infarction cases. However, because most patients requiring invasive treatment would have been transferred to higher-level centers, the degree of underestimation is likely limited.

Despite these limitations, the findings provide fundamental information for understanding the national emergency care system. Continuous surveillance using nationwide ED data remains essential for improving emergency preparedness, optimizing resource allocation, and enhancing patient outcomes, and will provide an important evidence base for future health policies and further research.

ARTICLE INFORMATION

Author contributions

Conceptualization: SA, JEK; Formal analysis: all authors; Investigation: all authors; Methodology: all authors; Supervision: SA, JEK; Writing–original draft: SA; Writing–review & editing: all authors. All authors read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

Funding

The authors received no financial support for this study.

Data availability

Data analyzed in this study were obtained from the National Emergency Medical Center (NEMC) under the Korean Ministry of Health and Welfare, and were used under license for the current study. Although the data are not publicly accessible, they are

available from the corresponding author upon reasonable request with permission from the NEMC.

REFERENCES

1. Greif R, Lauridsen KG, Djärv T, et al. European Resuscitation Council Guidelines 2025 executive summary. Resuscitation 2025;215 Suppl 1:110770.
2. Del Rios M, Bartos JA, Panchal AR, et al. Part 1: executive summary: 2025 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation 2025;152:S284–312.
3. Rao SV, O'Donoghue ML, Ruel M, et al. 2025 ACC/AHA/ACEP/NAEMSP/SCAI guideline for the management of patients with acute coronary syndromes: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 2025;151:e771–862.
4. The Lancet Regional Health-Western Pacific. Junior doctor strikes in South Korea: more doctors are needed? Lancet Reg Health West Pac 2024;44:101056.
5. Shin TG, Ko E, Han SH, Kim T, Choi DH. Epidemiology of sepsis in emergency departments: insights from the National Emergency Department Information System (NEDIS) database in Korea, 2018-2022. Clin Exp Emerg Med 2025;12:185–7.
6. Baldi E, Wnent J, Caputo ML, et al. European Resuscitation Council Guidelines 2025 epidemiology in resuscitation. Resuscitation 2025;215 Suppl 1:110733.
7. Park JS, Lee BK, Ko SK, Ro YS. Recent status of sudden cardiac arrests in emergency medical facilities: a report from the National Emergency Department Information System (NEDIS) of Korea, 2018-2022. Clin Exp Emerg Med 2023;10(S): S36–41.
8. Korean Statistical Information Service (KOSIS). Korean Ministry of Data and Statistics; [cited 2025 Nov 23]. Available from: https://kosis.kr/
9. GBD 2021 Causes of Death Collaborators. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2024;403:2100–32.
10. Ahn S, Ko E, Ro YS. Acute myocardial infarction diagnosed in emergency departments: a report from the National Emergency Department Information System (NEDIS) of Korea, 2018-2022. Clin Exp Emerg Med 2023;10(S): S42–7.
11. de Koning ER, Boogers MJ, Beeres SL, Kramer ID, Dannenberg WJ, Schalij MJ. Managing hospital capacity: achievements and lessons from the COVID-19 pandemic. Prehosp Disaster Med 2022;37:600–8.
12. Bouillon-Minois JB, Raconnat J, Clinchamps M, Schmidt J, Dutheil F. Emergency department and overcrowding during COVID-19 outbreak; a letter to editor. Arch Acad Emerg Med 2021;9:e28.
13. Sung HK, Paik JH, Lee YJ, Kang S. Impact of the COVID-19 outbreak on emergency care utilization in patients with acute myocardial infarction: a nationwide population-based study. J Korean Med Sci 2021;36:e111.
14. Contreras A, McIver W, Brown T, et al. The impact of the COVID-19 pandemic on adult out-of-hospital cardiac arrest incidence, community response, and outcomes in England: an interrupted time series analysis. Resuscitation 2025;216:110823.

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):98-104
https://doi.org/10.15441/ceem.26.020

Commentary

Emergency department presentations for suicide and self-harm in Korea, 2020–2024: an epidemiological study using the National Emergency Department Information System (NEDIS) database

Yuri Choi[1], Jinwoo Jeong[1], Borami Lim[2], Myeong Il Cha[2]

[1]Department of Emergency Medicine, Dong-A University Hospital, Dong-A University College of Medicine, Busan, Korea
[2]National Emergency Medical Center, National Medical Center, Seoul, Korea

pISSN: 3022-1811
eISSN: 2383-4625

Received: January 20, 2026
Revised: February 13, 2026
Accepted: February 18, 2026

Correspondence to:
Yuri Choi
(yurichoi@dau.ac.kr)

Co-correspondence to:
Myeong Il Cha
(cami2222@nmc.or.kr)

INTRODUCTION

Suicide represents a major global public health crisis. In 2021 alone, approximately 727,000 people worldwide died by suicide, exceeding the combined number of deaths caused by malaria, HIV, breast cancer, war, and homicide [1]. This figure corresponds to 1.1% of all deaths globally, indicating that more than one in every 100 deaths was attributable to suicide. Reflecting the severity of this issue, the reduction of suicide mortality has been incorporated as a key indicator within the United Nations Sustainable Development Goals, underscoring suicide prevention as a critical priority for the international community.

The manifestations of this crisis are also clearly observed in emergency care settings. In the United States, emergency department (ED) visits related to suicide attempts and intentional self-harm markedly increased from 1.43 million in 2011 to 5.37 million in 2021, imposing a substantial burden on the emergency care system [2]. The ED serves as a primary point of initial contact with the healthcare system for individuals who attempt suicide; therefore, it plays a pivotal role in suicide prevention strategies and initiatives to reduce suicide-related mortality.

The situation in Korea is even more concerning. Korea has consistently recorded the highest suicide rate among member countries of the Organization for Economic Co-operation and Development (OECD), a pattern that has persisted for more than a decade [3]. The COVID-19 pandemic has been associated with worsening mental health conditions and an increased risk of suicide [4], and instability within the healthcare system following the government's 2024 policy to expand medical school enrollment, accompanied by the withdrawal of resident physicians, has placed additional strain on the overall capacity of emergency medical services [5]. Such structural challenges within the healthcare system are likely to compromise the ability to respond adequately to psychiatric emergencies, including suicide and self-harm.

Although a previous study analyzed changes in ED utilization related to suicide and self-harm from 2018 to 2022, the time-dependent nature of suicide epidemiology necessitates continuous monitoring and reassessment of these trends. In particular, 2023, corresponding to the post–

How to cite this article:
Choi Y, Jeong J, Lim B, Cha MI. Emergency department presentations for suicide and self-harm in Korea, 2020–2024: an epidemiological study using the National Emergency Department Information System (NEDIS) database.
Clin Exp Emerg Med 2026;13(1):98-104.
https://doi.org/10.15441/ceem.26.020

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

© 2026 The Korean Society of Emergency Medicine

COVID-19 recovery phase, and 2024, during which structural changes in the emergency care system became pronounced due to healthcare workforce attrition [5], may have been characterized by distinct patterns of ED utilization and clinical outcomes among patients presenting with suicide attempts and self-harm compared with those observed in earlier periods. Accordingly, the present study extends the analysis period to more precisely evaluate the impact of recent environmental and healthcare system changes on suicide attempt– and self-harm–related indicators in ED settings.

METHODS

Ethics statement

This study was approved by the Institutional Review Board of the National Medical Center, Korea (No. NMC-2023-08-094). The requirement for informed consent was waived due to the retrospective, observational, and anonymous nature of the study.

Study population

This retrospective study used data from Korea's National Emergency Department Information System (NEDIS). NEDIS is a nationwide ED registry that has been managed by the National Emergency Medical Center since 2003 and contains administrative, clinical, and diagnostic information for all patients presenting to EDs. We analyzed patients who visited EDs nationwide between 2020 and 2024.

Patients with self-harm or suicide attempts were identified based on intentionality classification at ED presentation, specific initial triage classification codes recorded in the Korean Triage and Acuity Scale (KTAS), or discharge diagnoses using Korean Classification of Diseases (KCD) codes X60–X84 (excluding X65). The initial KTAS classification codes refer not to KTAS levels (1–5) but to standardized five-character alphanumeric codes assigned during the initial triage process. Patients were classified as having self-harm or suicide attempts if their initial triage code corresponded to one of the following categories: ABECA, BBECA, ABECB, BBECB, ABECD, BBECD, AABCC, BABCC, AABCD, or BABCD.

KCD code X65 (intentional self-poisoning by and exposure to alcohol) was excluded because alcohol-related poisoning recorded in ED administrative data is frequently subject to misclassification as accidental intoxication and may lack specificity for suicidal intent. Exclusion of X65 was intended to improve the specificity of case identification.

Patients who were dead on arrival were excluded. To evaluate age-specific characteristics, patients were categorized as pediatric (< 18 years), adult (18–64 years), or older adult (≥ 65 years).

Statistical analysis

Age- and sex-standardized ED visit rates per 100,000 population were calculated using the 2020 mid-year estimated population of Korea as the standard population. Age- and sex-standardized mortality rates per 100,000 ED visits were calculated using the number of ED visits recorded in NEDIS from 2020 to 2024 as the standard population. Temporal trends were assessed using linear regression models with calendar year included as a continuous variable, and P-values for trend were reported. Standardized incidence and mortality rates in 2023 and 2024 were compared using the z-test, and visit-based in-hospital mortality was compared using the chi-square test. All statistical analyses were performed using SAS ver. 9.4 (SAS Institute Inc) with a two-sided significance level of 0.05.

KEY OBSERVATIONS FROM NEDIS DATA

A total of 209,746 ED visits related to self-harm and suicide attempts were included in the analysis (Table 1). Overall, women accounted for a higher proportion of visits than men (63.2% vs. 36.8%). The proportion of women exceeded the overall average in the pediatric and adult groups, whereas men predominated in the older adult group. Most patients were treated at higher-level emergency medical centers, including regional emergency medical centers (42.6%) and local emergency medical centers (54.0%), and direct visits were more common than interhospital transfers (94.1% vs. 5.4%). Transport by the National Fire Agency ambulance service (119 ambulance) was the most frequent mode of arrival (64.0%).

The median ED length of stay was 3.7 hours (interquartile range [IQR], 1.9–7.7 hours), with the longest length of stay observed in the older adult group at 4.6 hours (IQR, 2.4–10.3 hours). The overall hospital admission rate was 29.1%, and more than half of older adult patients (51.2%) required hospitalization. Notably, the proportion of intensive care unit (ICU) admissions among patients with self-harm or suicide attempts was 15.5%, substantially exceeding the ICU admission rate among all ED patients (3.1%) [6].

Five-year trends in age- and sex-standardized ED visit rates, mortality rates, and visit-based in-hospital mortality stratified by age group are presented in Fig. 1. Age- and sex-standardized ED visit rates fluctuated year to year in the overall population and across all age groups; however, no statistically significant linear

Table 1. Baseline characteristics and outcomes of patients presenting with suicide attempt or self-harm

Characteristic	Total (n = 209,746)	Pediatric group (< 18 yr) (n = 22,640)	Adult group (18–64 yr) (n = 163,296)	Older adult group (≥ 65 yr) (n = 23,808)
Age (yr)	38.1 ± 19.2	15.3 ± 1.7	35.7 ± 13.4	76.0 ± 7.5
Sex				
Male	77,191 (36.8)	4,780 (21.1)	59,419 (36.4)	12,991 (54.6)
Female	132,555 (63.2)	17,860 (78.9)	103,877 (63.6)	10,817 (45.4)
Time from the symptom onset to ED arrival[a]	1.6 (0.9–4.6)	2.2 (1.0–8.8)	1.5 (0.9–4.1)	1.6 (0.9–4.4)
Type of ED				
Level I	89,436 (42.6)	9,978 (44.1)	68,326 (41.8)	11,131 (46.8)
Level II	113,314 (54.0)	12,043 (53.2)	89,446 (54.8)	11,825 (49.7)
Level III	6,996 (3.3)	619 (2.7)	5,524 (3.4)	852 (3.6)
Route of arrival				
Direct visit	197,459 (94.1)	20,894 (92.3)	154,743 (94.8)	21,822 (91.7)
Transfer from other hospital	11,328 (5.4)	1,540 (6.8)	7,856 (4.8)	1,931 (8.1)
Referral from outpatient clinic	882 (0.4)	200 (0.9)	634 (0.4)	48 (0.2)
Other	30 (0.0)	2 (0.0)	23 (0.0)	5 (0.0)
Unknown	47 (0.0)	4 (0.0)	40 (0.0)	2 (0.0)
Transport				
119 Ambulance	134,231 (64.0)	8,115 (35.8)	106,724 (65.4)	19,392 (81.5)
Other medical institution ambulance	1,173 (0.6)	91 (0.4)	811 (0.5)	271 (1.1)
Other ambulance	5,469 (2.6)	321 (1.4)	3,827 (2.3)	1,320 (5.5)
Police or official transport	1,929 (0.9)	143 (0.6)	1,713 (1.0)	73 (0.3)
Air transport	223 (0.1)	19 (0.1)	155 (0.1)	49 (0.2)
Other transport	65,467 (31.2)	13,712 (60.6)	49,097 (30.1)	2,658 (11.2)
Walk-in	1,125 (0.5)	220 (1.0)	868 (0.5)	37 (0.2)
Other/unknown	129 (0.1)	19 (0.1)	101 (0.1)	8 (0.0)
Length of stay (hr)[b]				
Mean ± SD	6.5 ± 8.6	5.7 ± 7.0	6.2 ± 8.1	8.7 ± 12.0
Median (IQR)	3.7 (1.9–7.7)	3.4 (1.8–6.6)	3.6 (1.8–7.6)	4.6 (2.4–10.3)
0–6	141,725 (67.6)	16,300 (72.0)	111,209 (68.1)	14,216 (59.7)
6–12	37,548 (17.9)	3,550 (15.7)	29,535 (18.1)	4,462 (18.7)
12–24	23,159 (11.0)	2,208 (9.8)	17,500 (10.7)	3,451 (14.5)
≥ 24	7,233 (3.4)	577 (2.5)	4,985 (3.1)	1,671 (7.0)
Unknown	81 (0.0)	5 (0.0)	67 (0.0)	8 (0.0)
ED disposition				
Discharge	133,008 (63.4)	15,883 (70.2)	108,782 (66.6)	8,343 (35.0)
Admission[c]	61,000 (29.1)	5,758 (25.4)	43,048 (26.4)	12,193 (51.2)
General ward	28,348 (13.5)	4,205 (18.6)	20,162 (12.3)	3,980 (16.7)
Intensive care unit	32,547 (15.5)	1,520 (6.7)	22,822 (14.0)	8,205 (34.5)
Transfer	8,267 (3.9)	717 (3.2)	6,108 (3.7)	1,442 (6.1)
Comfort care discharge	41 (0.0)	2 (0.0)	26 (0.0)	13 (0.1)
Death	5,458 (2.6)	203 (0.9)	3,519 (2.2)	1,736 (7.3)
Other/unknown	1,972 (0.9)	77 (0.3)	1,813 (1.1)	81 (0.3)
Hospital disposition				
Discharge	185,430 (88.4)	21,281 (94.0)	147,017 (90.0)	17,131 (72.0)
Transfer	13,369 (6.4)	971 (4.3)	8,948 (5.5)	3,450 (14.5)
Comfort care discharge	67 (0.0)	3 (0.0)	42 (0.0)	22 (0.1)
Death	8,129 (3.9)	254 (1.1)	4,897 (3.0)	2,978 (12.5)
Other/unknown	2,751 (1.3)	131 (0.6)	2,392 (1.5)	227 (1.0)

Values are presented as mean±SD, number (%), or median (IQR). Differences between totals and subgroup sums are due to missing age data (n=2); percentages may not total 100 due to rounding.

ED, emergency department; SD, standard deviation; IQR, interquartile range.

[a]Reporting was not obligatory for level III EDs. [b]Any ED length of stay over 5 days was treated as 5 days. [c]Admission locations other than general wards or intensive care units were unspecified.

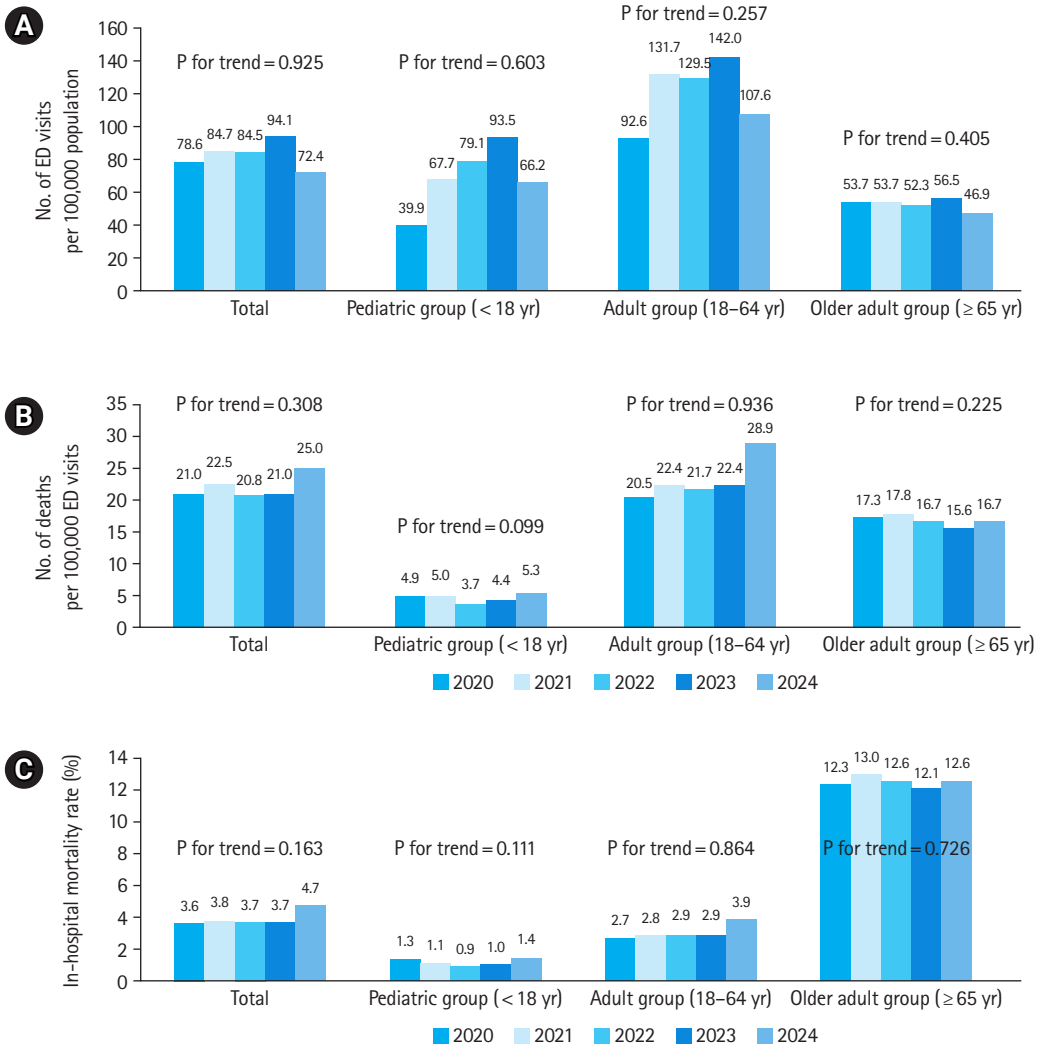


Fig. 1. Five-year trends. (A) Age- and sex-standardized emergency department (ED) visit rates per 100,000 population. (B) Age- and sex-standardized mortality rates per 100,000 ED visits. (C) In-hospital mortality among patients with suicide attempt or self-harm. P for trend was calculated using linear regression.

trend was observed over the 5-year period (P for trend > 0.05). Nevertheless, nonlinear patterns were noted in the adult and pediatric groups, with increases after 2021 followed by a decline in 2024.

Age- and sex-standardized mortality rates in the overall population showed a gradual upward tendency after 2020, although the 5-year linear trend did not reach statistical significance (P for trend = 0.308). In the adult group, an increasing trend in mortality was observed (P for trend = 0.099), whereas no clear temporal changes were identified in the pediatric or older adult groups. Visit-based in-hospital mortality also increased in 2024 compared with 2020 in the overall population; however, the linear trend across years was not statistically significant (P for trend = 0.163).

To evaluate the impact of healthcare system disruption, outcomes in 2023 and 2024 were compared (Table 2). Age- and sex-standardized ED visit rates decreased significantly in 2024 compared with 2023 in the overall population and across all age groups (z-test, all P < 0.05). In contrast, age- and sex-standardized mortality rates increased significantly in 2024 in the overall population and the adult group (all P < 0.05), whereas no significant differences were observed in the pediatric or older adult groups.

Visit-based in-hospital mortality also increased significantly in 2024 in the overall population and the adult group (Table 3). No statistically significant differences between years were observed in the pediatric or older adult groups.

Table 2. Comparison of age- and sex-standardized visit and mortality rates per 100,000 population among patients with suicide attempt or self-harm in 2023 and 2024

Variable	2023		2024		z	P-value
	Rate	SE	Rate	SE		
Incidence rate (per 100,000 population)						
Total	94.11	0.43	72.38	0.38	37.60	<0.05
Pediatric group (<18 yr)	93.49	1.14	66.19	0.96	18.37	<0.05
Adult group (18–64 yr)	141.96	0.98	107.57	0.85	26.57	<0.05
Older adult group (≥65 yr)	56.45	0.77	46.87	0.69	9.29	<0.05
Mortality rate (per 100,000 ED visits)						
Total	21.03	0.50	24.98	0.61	−5.01	<0.05
Pediatric group (<18 yr)	4.36	0.55	5.34	0.69	−1.10	0.27
Adult group (18–64 yr)	22.42	0.70	28.89	0.89	−5.74	<0.05
Older adult group (≥65 yr)	15.64	1.08	16.70	1.14	−0.67	0.501

Comparisons between 2023 and 2024 were performed using the z-test. A two-sided P<0.05 was considered statistically significant.
SE, standard error; ED, emergency department.

Table 3. Comparison of visit-based in-hospital mortality among patients with suicide attempt or self-harm in 2023 and 2024

Group	Visit-based in-hospital mortality (%)		P-value
	2023	2024	
Total	3.70	4.72	<0.05
Pediatric group (<18 yr)	1.04	1.42	0.104
Adult group (18–64 yr)	2.87	3.89	<0.05
Older adult group (≥65 yr)	12.13	12.55	0.540

Visit-based in-hospital mortality was defined as the proportion of in-hospital deaths among patients visiting the emergency department. Comparisons between 2023 and 2024 were performed using the chi-square test, with a two-sided significance level of 0.05.

INTERPRETATION AND IMPLICATIONS

This study analyzed the characteristics and temporal trends of patients presenting with self-harm and suicide attempts between 2020 and 2024 using data from NEDIS. During this period, Korea experienced a series of unprecedented healthcare crises, including the COVID-19 pandemic in 2020, a subsequent recovery phase from 2021 to 2023, and a major disruption to the healthcare system in 2024 related to the medical workforce crisis. By examining long-term changes in the incidence and outcomes of self-harm– and suicide attempt–related ED visits across these overlapping crises, this study provides meaningful insights into the epidemiology of psychiatric emergencies under sustained healthcare system stress.

Age- and sex-standardized ED visit rates increased gradually after 2020 and then declined sharply in 2024, a pattern that was particularly pronounced in the adult age group. In a previous study analyzing ED data from a tertiary hospital in Incheon during the COVID-19 period, COVID-19 itself was identified as an independent risk factor for suicide attempts, and statistically significant increases were observed in suicide attempts with lower rescue potential as well as shifts in socioeconomic motivations among patients presenting after suicide attempts [4]. These findings suggest that large-scale social crises, such as infectious disease outbreaks, may amplify suicide risk, which is consistent with the increasing incidence observed after 2020 in the present study.

When 5-year incidence trends were examined by age group, incidence rates in the older adult group remained relatively stable over time. In contrast, both the adult and pediatric groups demonstrated a noticeable increase in age- and sex-standardized incidence rates from 2020 to 2022, followed by a decline in 2024. Although the overall 5-year linear trend did not reach statistical significance, the upward pattern during the pandemic and recovery period was evident in these younger age groups. These findings are consistent with those of several previous studies comparing the pre-pandemic and pandemic periods [4,7]. Notably, one of the most striking findings of the present study was the pronounced increase in self-harm and suicide attempts among adolescents. This subgroup was estimated to be predominantly composed of adolescents, as the mean age was 15.3 ± 1.7 years. This aligns with the World Health Organization statistics indicating that suicide ranked as the third leading cause of death among individuals aged 15–29 years in 2021, after road traffic injuries and interpersonal violence [8]. In addition, a prior study analyzing nationwide ED data in Korea from 2016 to 2020 reported a sustained increase in ED visit rates due to suicide attempts and self-harm among young adults aged 19–29 years [3]. Increasing incidence of suicide attempts and self-harm among adolescents and young adults has been reported globally [9], highlighting the need for more proactive and targeted policy interventions focused on these age groups.

In the older adult group, incidence rates did not change substantially over time; however, hospitalization rates, ICU admission

rates, and mortality were markedly higher than in other age groups. The hospitalization rate among older adults was 51.2%, substantially exceeding that of the overall study population, and the proportion of ICU admissions was also considerably higher than that observed among all ED patients. Age- and sex-standardized mortality rates were highest in the older adult group, consistent with previous studies reporting a substantially higher likelihood of death following suicide attempts among older individuals than among younger age groups [3]. Suicide attempts in older adults are more likely to involve planned behaviors associated with physical suffering or social isolation and are therefore more frequently fatal [10]. In addition, older patients often have multiple medical and surgical comorbidities [11], which may exacerbate the clinical consequences of self-harm or suicide attempts and contribute to mortality from causes beyond the index event [12,13]. These findings suggest the need for policy approaches tailored specifically to older adults, distinct from those addressing the increasing trends observed among adolescents.

In the context of these increasing trends, the sharp decline in incidence observed in 2024 is difficult to interpret as a true reduction in risk. Instead, it may reflect reduced access to emergency medical services during the healthcare system disruption, such that individuals requiring urgent care for self-harm or suicide attempts were less likely to be captured within the existing emergency care safety net.

In addition, analyses of fire service transport records have shown that a substantial proportion of emergency patients, including those with potentially severe conditions, are transported to facilities outside designated emergency medical center level institutions and that regional transport patterns differ considerably from national averages [14]. These findings suggest that administrative ED databases such as NEDIS may not fully capture all emergency transports at the national level. Accordingly, the observed decrease in incidence may not indicate a true decline in occurrence but may instead reflect that a subset of patients was not represented in the available data.

Meanwhile, although most patients with self-harm or suicide attempts in this study were treated at center-level emergency medical facilities, this finding may partially reflect limitations of the NEDIS data. In some institutions, information on intentionality may not have been transmitted, potentially resulting in undercapture of eligible patients, and it is unlikely that all individuals with self-harm or suicide attempts entered the emergency care system through NEDIS-participating institutions. In particular, during periods such as the COVID-19 pandemic or the healthcare system disruption, patients may have been redistributed to non-participating facilities, including outpatient clinics or emergency care posts. Under these circumstances, patients who did present to NEDIS-reporting institutions may have represented a relatively higher-acuity population, which could provide one plausible explanation for the concurrent observation of decreased incidence and increased mortality in this study.

The increase in standardized mortality observed in 2024 may suggest excess mortality; however, this interpretation warrants caution given the concurrent decrease in age- and sex-standardized incidence rates. This pattern raises the possibility that patients with less severe self-harm or suicide attempts were not captured in the ED data. Such lower-acuity patients may have been diverted to healthcare settings outside the emergency care system, thereby reducing the number of ED presentations and increasing the overall severity among patients who did present to EDs, which in turn could manifest as an apparent increase in mortality. To determine more clearly whether excess mortality occurred, further studies incorporating data sources outside the emergency medical system, such as national fire service dispatch records, will be required.

Finally, the findings of this study suggest that many longitudinal investigations limited to ED populations may not fully represent the broader population of individuals who die by suicide. Notably, fewer than 16% of individuals who attempt suicide are reported to receive care from mental health professionals, and this proportion has been shown to decline by an average of 9.1% annually [2]. These findings indicate that hospital-based interventions alone are insufficient for effective suicide prevention and underscore the importance of strategies aimed at early identification and sustained support of high-risk individuals in the community who remain disconnected from the emergency medical care system.

ARTICLE INFORMATION

Author contributions

Conceptualization: YC, JJ; Formal analysis: all authors; Investigation: all authors; Methodology: all authors; Supervision: JJ, MIC; Writing–original draft: YC; Writing–review & editing: all authors. All authors read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

Funding

The authors received no financial support for this study.

Data availability

Data analyzed in this study were obtained from the National Emergency Medical Center (NEMC) under the Korean Ministry of Health and Welfare, and were used under license for the current study. Although the data are not publicly accessible, they are available from the corresponding author upon reasonable request with permission from the NEMC.

REFERENCES

1. World Health Organization. Suicide worldwide in 2021: global health estimates. World Health Organization; 2025.
2. Bommersbach TJ, Olfson M, Rhee TG. National trends in emergency department visits for suicide attempts and intentional self-harm. Am J Psychiatry 2024;181:741–52.
3. Lee KS, Sung HK, Yoo SY, Min HS. Temporal trends and characteristics of adult patients in emergency department related to suicide attempt or self-harm in Korea, 2016-2020. J Korean Med Sci 2023;38:e40.
4. Kim H, Durey A, Kang S, et al. Factors affecting patients who attempted suicide in the emergency department due to the prolonged pandemic of COVID-19. Clin Exp Emerg Med 2023;10:418–25.
5. Yoon JH, Kwon IH, Park HW. The South Korean health-care system in crisis. Lancet 2024;403:2589.
6. Park HA, Kim T, Kim HJ, Han SH. Epidemiological trends in emergency department visits by age group: a report from the National Emergency Department Information System (NEDIS) of Korea, 2020-2024. Clin Exp Emerg Med 2025;12:405–13.
7. Lee J, Kim D, Lee WJ, Woo SH, Jeong S, Kim SH. Association of the COVID-19 pandemic and low-rescue suicide attempts in patients visiting the emergency department after attempting suicide. J Korean Med Sci 2021;36:e243.
8. Knipe D, Padmanathan P, Newton-Howes G, Chan LF, Kapur N. Suicide and self-harm. Lancet 2022;399:1903–16.
9. Hawton K, Bale L, Brand F, et al. Mortality in children and adolescents following presentation to hospital after non-fatal self-harm in the Multicentre Study of Self-harm: a prospective observational cohort study. Lancet Child Adolesc Health 2020;4:111–20.
10. Jung KY, Kim T, Cha WC. Self-harm characteristics of younger-old and older-old adults admitted to emergency departments: a nationwide study. Signa Vitae 2023;19:159–66.
11. Kim S, Kang H, Cho Y, et al. Emergency department utilization and risk factors for mortality in older patients: an analysis of Korean National Emergency Department Information System data. Clin Exp Emerg Med 2021;8:128–36.
12. Goldman-Mellor S, Olfson M, Lidon-Moyano C, Schoenbaum M. Association of suicide and other mortality with emergency department presentation. JAMA Netw Open 2019;2:e1917571.
13. Goldman-Mellor S, Olfson M, Schoenbaum M. Trends in mortality after emergency department presentation for suicidal behaviour in California. Inj Prev 2025;31:594–6.
14. Park J, Jung J, Seo B, et al. [Analysis and policy recommendations for problems in the emergency transport system]. Research Report No. 2024-17. Korean Medical Association Research Institute for Healthcare Policy; 2025.

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):105-109
https://doi.org/10.15441/ceem.25.206

Commentary

Emergency medicine physicians treating urine cultures when it is not clear they should

Amanda Doodlesack[1], Clifford Marks[1], Nicolette Castaneda[2], Jonathan Edlow[1], Shamai A. Grossman[1]

[1]Department of Emergency Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA
[2]Department of Emergency Medicine, CarePoint Health, Greenwood Village, CO, USA

pISSN: 3022-1811
eISSN: 2383-4625

Received: September 5, 2025
Revised: November 27, 2025
Accepted: November 27, 2025

Correspondence to:
Amanda Doodlesack
(adoodles@bidmc.harvard.edu)

Emergency physicians make treatment decisions about positive urine cultures that result after patients are discharged from the emergency department (ED). The Infectious Diseases Society of America (IDSA) incorporates both symptoms and laboratory findings into the diagnosis of a urinary tract infection (UTI). Typical symptoms include dysuria, urinary frequency, urgency, suprapubic pain or tenderness, and flank pain, while confirmation of a UTI requires a positive urine culture with ≥ 100,000 colony-forming units (CFU) per milliliter of a single organism [1]. Although many cases involve patients who presented with clear symptoms of a UTI, a substantial proportion involve atypical presentations (e.g., confusion or a fall in an older patient) or patients without symptoms plausibly attributable to a UTI. Such positive culture results are considered asymptomatic bacteriuria (ASB) [1].

The IDSA recommends against antibiotic treatment for nearly all cases of asymptomatic bacteriuria other than in pregnant patients. Despite these recommendations, prescribers treat an average of 45% of ASB cases with antibiotics, according to a recent meta-analysis [2]. However, this and a more recent study addressing ASB in the emergency setting use conservative definitions, excluding cases with vague symptoms unless those symptoms are clearly attributable to another cause [2,3]. These scenarios, which emergency physicians frequently face, have been understudied. Our earlier paper examining the binary decision to treat with antibiotics demonstrated frequent deviation from IDSA recommendations both in these scenarios and in cases that fit a conservative ASB definition [4]. As a follow-up, this study aims to clarify how emergency physicians manage ASB by examining antibiotic selection and treatment duration following a positive culture result in ED patients who are asymptomatic or present with symptoms atypical for UTI.

We surveyed emergency medicine physicians based at an academic tertiary care medical center and its affiliated hospital network, which includes six community hospitals. The study used a cross-sectional design. An electronic questionnaire was emailed to all emergency physicians across the network and addressed nine common scenarios encountered when following up a positive urine culture, including several involving asymptomatic or minimally symptomatic bacteriuria. These scenarios were selected based on existing IDSA guidelines and to reflect the range of emergency practice [1,5–10]. ASB was defined as one or more bacterial species growing in the urine at ≥ 100,000 CFU/mL in the absence of signs or symptoms of a UTI [1]. Responses were collected over a 3-month period.

How to cite this article:
Doodlesack A, Marks C, Castaneda N, Edlow J, Grossman SA. Emergency medicine physicians treating urine cultures when it is not clear they should.
Clin Exp Emerg Med 2026;13(1):105-109.
https://doi.org/10.15441/ceem.25.206

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

© 2026 The Korean Society of Emergency Medicine

The survey was distributed to 305 emergency physicians, with 120 responses received, yielding a response rate of 39.3%. Of these respondents, 49 (40.8%) worked in academic medical centers with affiliated emergency medicine residency programs, 45 (37.5%) in large community hospitals (greater than 40,000 visits per year), and 26 (21.7%) in small community hospitals (less than 40,000 visits per year). Survey participants included 14 residents (11.7%), 33 attendings (27.5%) with less than 5 years out of training, 23 attendings (19.2%) with 5–9 years out, 37 attendings (30.8%) with 10–20 years out, and 13 attendings (10.8%) with more than 20 years out (Table 1).

Cephalexin was the most commonly prescribed antibiotic (selected 344 times, 46.9%), followed by cefpodoxime (162 times, 22.1%) and nitrofurantoin (144 times, 19.6%) (Table 2, Fig. 1). The only scenarios in which cephalexin was not the most common choice were the asymptomatic female patient (scenario 1), in which nitrofurantoin was most frequently selected, and the asymptomatic post–kidney transplant male patient (scenario 9), in which cefpodoxime was most common. Overall, ciprofloxacin was the least commonly prescribed agent. In five of the nine scenarios, none of the 120 emergency physicians selected ciprofloxacin. The only scenarios in which it was not the least commonly chosen antibiotic were the asymptomatic male patients (scenarios 3 and 7) and the asymptomatic post–kidney transplant patient (scenario 9).

Table 1. Demographic characteristics of the respondents (n=120)

Characteristic	No. of respondents (%)
Practice setting	
Academic medical center	49 (40.8)
Large community hospital	45 (37.5)
Small community hospital	26 (21.7)
No. of years out of training	
Resident	14 (11.7)
< 5	33 (27.5)
5–9	23 (19.2)
10–20	37 (30.8)
> 20	13 (10.8)

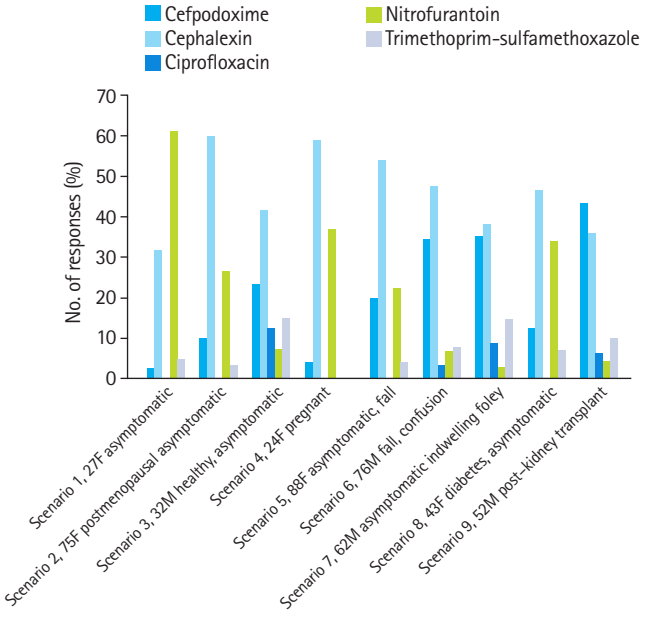


Fig. 1. Antibiotic selection by clinical scenario: (1) a 27-year-old asymptomatic nonpregnant woman; (2) a 75-year-old asymptomatic postmenopausal woman; (3) a 32-year-old asymptomatic man; (4) a 24-year-old asymptomatic pregnant woman; (5) an 88-year-old asymptomatic woman after a fall without urinary symptoms; (6) a 76-year-old man with confusion after a fall without urinary symptoms; (7) a 62-year-old asymptomatic man with an indwelling Foley catheter; (8) a 43-year-old asymptomatic woman with diabetes; and (9) a 52-year-old asymptomatic man post–kidney transplant.

Table 2. Number of respondents selecting each antibiotic by scenario

Clinical scenario	Cefpodoxime	Cephalexin	Ciprofloxacin	Nitrofurantoin	Trimethoprim-sulfamethoxazole
1 (n = 41)	1 (2.4)	13 (31.7)	0 (0)	25 (61.0)	2 (4.9)
2 (n = 60)	6 (10.0)	36 (60.0)	0 (0)	16 (26.7)	2 (3.3)
3 (n = 120)	28 (23.3)	50 (41.7)	15 (12.5)	9 (7.5)	18 (15.0)
4 (n = 119)	5 (4.2)	70 (58.8)	0 (0)	44 (37.0)	0 (0)
5 (n = 76)	15 (19.7)	41 (53.9)	0 (0)	17 (22.4)	3 (3.9)
6 (n = 116)	40 (34.5)	55 (47.4)	4 (3.4)	8 (6.9)	9 (7.8)
7 (n = 34)	12 (35.3)	13 (38.2)	3 (8.8)	1 (2.9)	5 (14.7)
8 (n = 56)	7 (12.5)	26 (46.4)	0 (0)	19 (33.9)	4 (7.1)
9 (n = 111)	48 (43.2)	40 (36.0)	7 (6.3)	5 (4.5)	11 (9.9)
All scenarios (n = 733)	162 (22.1)	344 (46.9)	29 (4.0)	144 (19.6)	54 (7.4)

Values are presented as number (%). Percentages may not total 100 due to rounding. Clinical scenarios are as follows: (1) a 27-year-old asymptomatic nonpregnant woman; (2) a 75-year-old asymptomatic postmenopausal woman; (3) a 32-year-old asymptomatic man; (4) a 24-year-old asymptomatic pregnant woman; (5) an 88-year-old asymptomatic woman after a fall without urinary symptoms; (6) a 76-year-old man with confusion after a fall without urinary symptoms; (7) a 62-year-old asymptomatic man with an indwelling Foley catheter; (8) a 43-year-old asymptomatic woman with diabetes; and (9) a 52-year-old asymptomatic man post–kidney transplant.

Overall, a 7-day course of antibiotics was most commonly selected (338 times, 46.2%), followed by 5 days (240 times, 32.8%) and 10 days (122 times, 16.7%) (Table 3, Fig. 2). A 3-day course was rarely selected across all scenarios. Clinical scenarios in which a 5-day course was more frequently chosen included the 27-year-old asymptomatic woman (scenario 1), the 75-year-old postmenopausal asymptomatic woman (scenario 2), the 24-year-old pregnant woman (scenario 4), and the 43-year-old asymptomatic woman with diabetes (scenario 8). A 10-day course was selected more often in all scenarios involving a male patient, including the 32-year-old asymptomatic man (scenario 3), the 76-year-old man with confusion after a fall (scenario 6), the 62-year-old asymptomatic man with an indwelling Foley catheter (scenario 7), and the 52-year-old man post–kidney transplant (scenario 9).

Table 3. Number of respondents selecting each duration of treatment by clinical scenario

Clinical scenario	Duration of treatment (day)			
	3	5	7	10
1 (n = 41)	9 (22.0)	25 (61.0)	7 (17.1)	0 (0)
2 (n = 60)	5 (8.3)	31 (51.7)	24 (40.0)	0 (0)
3 (n = 119)	2 (1.7)	16 (13.4)	60 (50.4)	41 (34.5)
4 (n = 119)	8 (6.7)	54 (45.4)	53 (44.5)	4 (3.4)
5 (n = 76)	2 (2.6)	37 (48.7)	37 (48.7)	0 (0)
6 (n = 116)	0 (0)	22 (19.0)	65 (56.0)	29 (25.0)
7 (n = 34)	0 (0)	3 (8.8)	18 (52.9)	13 (38.2)
8 (n = 56)	2 (3.6)	40 (71.4)	13 (23.2)	1 (1.8)
9 (n = 111)	4 (3.6)	12 (10.8)	61 (55.0)	34 (30.6)
All scenarios (n = 732)	32 (4.4)	240 (32.8)	338 (46.2)	122 (16.7)

Values are presented as number (%). Percentages may not total 100 due to rounding. Clinical scenarios are as follows: (1) a 27-year-old asymptomatic nonpregnant woman; (2) a 75-year-old asymptomatic postmenopausal woman; (3) a 32-year-old asymptomatic man; (4) a 24-year-old asymptomatic pregnant woman; (5) an 88-year-old asymptomatic woman after a fall without urinary symptoms; (6) a 76-year-old man with confusion after a fall without urinary symptoms; (7) a 62-year-old asymptomatic man with an indwelling Foley catheter; (8) a 43-year-old asymptomatic woman with diabetes; and (9) a 52-year-old asymptomatic man post–kidney transplant.

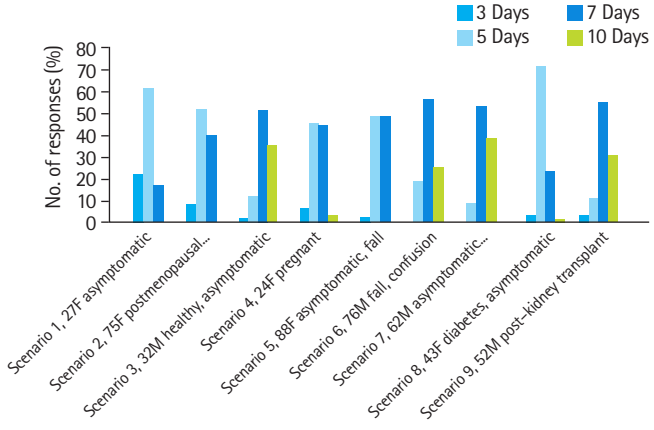


Fig. 2. Treatment duration by clinical scenario: (1) a 27-year-old asymptomatic nonpregnant woman; (2) a 75-year-old asymptomatic postmenopausal woman; (3) a 32-year-old asymptomatic man; (4) a 24-year-old asymptomatic pregnant woman; (5) an 88-year-old asymptomatic woman after a fall without urinary symptoms; (6) a 76-year-old man with confusion after a fall without urinary symptoms; (7) a 62-year-old asymptomatic man with an indwelling Foley catheter; (8) a 43-year-old asymptomatic woman with diabetes; and (9) a 52-year-old asymptomatic man post–kidney transplant.

The emergency physicians surveyed demonstrated considerable variability in both the type and duration of antibiotics prescribed. Across all scenarios, most physicians selected either a 5- or 7-day treatment course, and the most common antibiotic choices were cefpodoxime and cephalexin. In the single scenario with clear IDSA recommendations for treatment, ASB in a pregnant woman, physicians largely followed established guidelines; 89.9% (107 of 119) selected a 5- or 7-day antibiotic course, which corresponds to the IDSA recommendation of 4 to 7 days of therapy [1]. In this scenario, all physicians selected either nitrofurantoin or a β-lactam antibiotic (cefpodoxime or cephalexin), again matching IDSA guidance [1].

Variability in practice patterns appeared most pronounced in scenarios involving vague symptoms that could plausibly be related to a UTI. For example, for the 76-year-old man with confusion and a fall (scenario 6), respondents selected all five antibiotic options, and treatment durations ranged between 5 and 10 days, indicating no consensus. Similarly, multiple respondents treating ASB in a kidney transplant patient selected various antibiotics and durations, reflecting the limited evidence available to guide decisions in these cases [11]. The frequent selection of cephalexin was particularly noteworthy because it generally requires dosing four times daily, which can make adherence more difficult.

For ASB scenarios without symptoms suggestive of UTI, antibiotic selections often diverged from current IDSA guidelines for cystitis treatment. The latest IDSA recommendations for women with cystitis advise a 5-day course of nitrofurantoin or a 3-day course of trimethoprim-sulfamethoxazole as first-line therapy, reserving β-lactams for situations in which first-line agents cannot be used [12]. Although physicians predominantly chose nitrofurantoin when treating a young woman with bacteriuria and no symptoms, they tended to favor β-lactams when the patient had diabetes or was older.

Physicians generally prescribed longer antibiotic courses for male patients, consistent with the traditional view that male UTIs are "complicated" because of anatomic considerations. However,

recent studies challenge this assumption and suggest that 7-day courses may be adequate for men [13–15].

Ciprofloxacin and trimethoprim-sulfamethoxazole were rarely selected across all clinical scenarios. This pattern may reflect concerns about adverse events associated with fluoroquinolones and high community resistance rates to trimethoprim-sulfamethoxazole [16].

The substantial variability observed in this follow-up study highlights the challenges emergency physicians face when balancing UTI management with antibiotic stewardship. Because patients are often lost to follow-up, physicians may favor broader-spectrum or longer-duration therapy even when evidence suggests such practices are unnecessary. Although patients with classic UTI symptoms and positive cultures require treatment to prevent complications such as pyelonephritis or urosepsis, overtreatment of ASB persists as a significant problem [6]. Inappropriate antibiotic use contributes to resistance at both individual and community levels [1,7]. Emergency physicians must therefore weigh risks and benefits carefully, tailoring choices about antibiotic type and duration to each clinical situation.

Overtreatment partly arises from a culture of excessive urinalysis and culture ordering in ED settings, driven by concerns about missing occult infection. One study of an academic emergency department found that urinalyses were ordered for one quarter of patients presenting with chest pain or headache [17]. Efforts to reduce mistreatment include requiring a second order for a culture after an initial urinalysis, canceling low-risk culture orders, and limiting reflex culture practices [18–21]. Some departments have also incorporated pharmacist review of cultures to guide antibiotic choice and duration and to decrease unnecessary treatment. As a result, studies have shown improved adherence to IDSA guidelines and, in one case, successful discontinuation of antibiotics for ASB [22–24]. However, the extent to which such initiatives are widely implemented is unclear, and inappropriate treatment of ASB, as demonstrated in our studies, likely remains common.

Limitations of this study include its relatively small sample size from a single network of emergency physicians, which may reduce generalizability. Nonetheless, the sample encompassed a wide range of physicians, from residents to those with more than 20 years of experience, as well as individuals practicing in small community, large community, and academic hospitals, and physicians currently practicing in other healthcare settings. Although the survey response rate was modest, the diversity of the network likely provides a reasonable cross-section of emergency physicians. Response bias is also possible because participation was voluntary, and self-reported practices may not perfectly reflect actual behavior.

In conclusion, this follow-up study demonstrates substantial variation in emergency physicians' choices regarding antibiotic selection and duration. Most physicians treat positive urine cultures for 5 or 7 days and most frequently prescribe cephalexin. While timely treatment is necessary to prevent progression and complications, unnecessary antibiotic use can cause harm both to individual patients and to the broader community. Further research is needed to better inform treatment decisions for bacteriuria in patients with vague or atypical symptoms and to clarify the factors influencing management choices.

ARTICLE INFORMATION

Ethics statement

This study was exempted from an institutional review board approval because it was considered a quality assurance project, and participation was voluntary. All participants provided online informed consent before beginning the survey.

Author contributions

Conceptualization: SAG, JE; Data curation: AD, NC; Formal analysis: all authors; Investigation: SAG, JE, AD, NC; Methodology: SAG, JE, AD, NC; Project administration: SAG; Resources: SAG, JE; Software: NC, AD, SAG; Supervision: SAG, JE; Validation: SAG, JE; Visualization: SAG, JE, AD, CM; Writing–original draft: AD; Writing–review & editing: all authors. All authors read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

Funding

The authors received no financial support for this study.

Data availability

Data analyzed this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Nicolle LE, Gupta K, Bradley SF, et al. Clinical practice guideline for the management of asymptomatic bacteriuria: 2019 update by the Infectious Diseases Society of America. Clin Infect Dis 2019;68:e83–110.

2. Flokas ME, Andreatos N, Alevizakos M, Kalbasi A, Onur P, Mylonakis E. Inappropriate management of asymptomatic patients with positive urine cultures: a systematic review and meta-analysis. Open Forum Infect Dis 2017;4:ofx207.
3. Veillette JJ, Waters CD, Gelman SS, et al. Antibiotic prescribing for adult bacteriuria and pyuria in community hospital emergency departments. Am J Emerg Med 2021;40:1–5.
4. Castañeda NR, Doodlesack A, Fyfe W, Edlow J, Grossman SA. The data and the reality: urine cultures and emergency medicine physicians. Intern Emerg Med 2022;17:2349–55.
5. Trautner BW, Grigoryan L. Approach to a positive urine culture in a patient without urinary symptoms. Infect Dis Clin North Am 2014;28:15–31.
6. Gupta K. Acute simple cystitis in female adults. In: UpToDate [Internet]. Wolters Kluwer; [updated 2026 Jan 9; cited 2020 May 20]. Available from: https://www.uptodate.cn/contents/acute-simple-cystitis-in-female-adults
7. Mody L, Juthani-Mehta M. Urinary tract infections in older women: a clinical review. JAMA 2014;311:844–54.
8. Colgan R, Williams M. Diagnosis and treatment of acute uncomplicated cystitis. Am Fam Physician 2011;84:771–6.
9. Matuszkiewicz-Rowińska J, Małyszko J, Wieliczko M. Urinary tract infections in pregnancy: old and new unresolved diagnostic and therapeutic problems. Arch Med Sci 2015;11:67–77.
10. Cortes-Penfield NW, Trautner BW, Jump RL. Urinary tract infection and asymptomatic bacteriuria in older adults. Infect Dis Clin North Am 2017;31:673–88.
11. Goldman JD, Julian K. Urinary tract infections in solid organ transplant recipients: guidelines from the American Society of Transplantation Infectious Diseases Community of Practice. Clin Transplant 2019;33:e13507.
12. Gupta K, Hooton TM, Naber KG, et al. International clinical practice guidelines for the treatment of acute uncomplicated cystitis and pyelonephritis in women: a 2010 update by the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases. Clin Infect Dis 2011;52:e103–20.
13. Drekonja DM, Rector TS, Cutting A, Johnson JR. Urinary tract infection in male veterans: treatment patterns and outcomes. JAMA Intern Med 2013;173:62–8.
14. Drekonja DM, Trautner B, Amundson C, Kuskowski M, Johnson JR. Effect of 7 vs 14 days of antibiotic therapy on resolution of symptoms among afebrile men with urinary tract infection: a randomized clinical trial. JAMA 2021;326:324–31.
15. Germanos GJ, Trautner BW, Zoorob RJ, et al. No clinical benefit to treating male urinary tract infection longer than seven days: an outpatient database study. Open Forum Infect Dis 2019;6:ofz216.
16. Umarje SP, Alexander CG, Cohen AJ. Ambulatory fluoroquinolone use in the United States, 2015-2019. Open Forum Infect Dis 2021;8:ofab538.
17. Stagg A, Lutz H, Kirpalaney S, et al. Impact of two-step urine culture ordering in the emergency department: a time series analysis. BMJ Qual Saf 2018;27:140–7.
18. Patel HP. The abnormal urinalysis. Pediatr Clin North Am 2006;53:325–37.
19. Jones CW, Culbreath KD, Mehrotra A, Gilligan PH. Reflect urine culture cancellation in the emergency department. J Emerg Med 2014;46:71–6.
20. Munigala S, Rojek R, Wood H, et al. Effect of changing urine testing orderables and clinician order sets on inpatient urine culture testing: analysis from a large academic medical center. Infect Control Hosp Epidemiol 2019;40:281–6.
21. Randolph AG, Reder L, Englund JA. Risk of bacterial infection in previously healthy respiratory syncytial virus-infected young children admitted to the intensive care unit. Pediatr Infect Dis J 2004;23:990–4.
22. Giruzzi ME, Tawwater JC, Grelle JL. Evaluation of antibiotic utilization in an emergency department after implementation of an antimicrobial stewardship pharmacist culture review service. Hosp Pharm 2020;55:261–7.
23. Zhang X, Rowan N, Pflugeisen BM, Alajbegovic S. Urine culture guided antibiotic interventions: a pharmacist driven antimicrobial stewardship effort in the ED. Am J Emerg Med 2017;35:594–8.
24. Gupta K. Acute simple cystitis in male adults. In: UpToDate [Internet]. Wolters Kluwer; [updated 2025 Jan 23; cited 2023 May 20]. Available from: https://www.uptodate.cn/contents/acute-simple-cystitis-in-male-adults

Clinical and Experimental Emergency Medicine

Clin Exp Emerg Med 2026;13(1):110-111
https://doi.org/10.15441/ceem.25.011

Images in Emergency Medicine

A false-positive Focused Assessment with Sonography in Trauma (FAST) in a trauma patient

Lindsay Taylor[1], Fred Milgrim[2], Matthew Riscinti[2], Michael Gottlieb[3]

[1]Department of Emergency Medicine, Virginia Commonwealth University Medical Center, Richmond, VA, USA
[2]Department of Emergency Medicine, Denver Health and Hospital Authority, University of Colorado, Denver, CO, USA
[3]Department of Emergency Medicine, Rush University Medical Center, Chicago, IL, USA

pISSN: 3022-1811
eISSN: 2383-4625

Received: January 17, 2025
Revised: March 18, 2025
Accepted: March 19, 2025

Correspondence to:
Lindsay Taylor
(lataylor18@gmail.com)

A 26-year-old man presented to the emergency department after a dirt bike accident, reporting right rib pain and shortness of breath. He was tachycardic with otherwise normal vital signs. On examination, he had equal breath sounds and tenderness to the right chest wall. After the primary survey, a Focused Assessment with Sonography in Trauma (FAST) was performed with concern for free fluid in the right upper quadrant (Fig. 1A). He was deemed stable for computed tomography (CT) imaging to investigate for injuries (Fig. 1B). In this case, CT confirmed the ultrasound finding to be an artifact.

The FAST examination is an important tool in the assessment of trauma patients and is highly specific for hemoperitoneum [1,2]. In an unstable patient with abdominal trauma, a positive finding on the FAST will often guide the decision for operative management; stable patients will undergo CT imaging. A positive FAST examination is considered when there is an anechoic stripe present between the liver and kidney. A recent study described the "lipliner sign," an artifact creating a false anechoic stripe between the liver and kidney on the FAST caused by post-processing imaging enhancements that are now standard on many modern ultrasound machines [3]. This artifact can be reduced by adjusting the machine settings: (1) decrease speckle reduction; (2) increase the dynamic range; (3) increase dehaze; or (4) decrease the gray map. While such

Capsule Summary

What is already known

The focused assessment with sonography in trauma (FAST) exam is highly specific in detecting hemoperitoneum during early trauma evaluations and ruptured ectopic pregnancies, facilitating emergent surgery. Free fluid found in the hepatorenal recess is the most sensitive and earliest sign of hemoperitoneum.

What is new in the current study

Post-processing imaging enhancements are now standard on many modern ultrasound machines. These enhancements can cause a false-positive anechoic stripe between the liver and kidney on the FAST examination. This artifact can be reduced by decreasing speckle reduction, increasing the dynamic range, increasing dehaze, and decreasing the gray map. While these adjustments appear to decrease this artifact, there is a tradeoff in image quality. Without implementing these changes, awareness of this artifact is critical to avoid unnecessary computed tomography imaging or surgery.

How to cite this article:
Taylor L, Milgrim F, Riscinti M, Gottlieb M. A false-positive Focused Assessment with Sonography in Trauma (FAST) in a trauma patient.
Clin Exp Emerg Med 2026;13(1):110-111.
https://doi.org/10.15441/ceem.25.011

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/).

© 2026 The Korean Society of Emergency Medicine

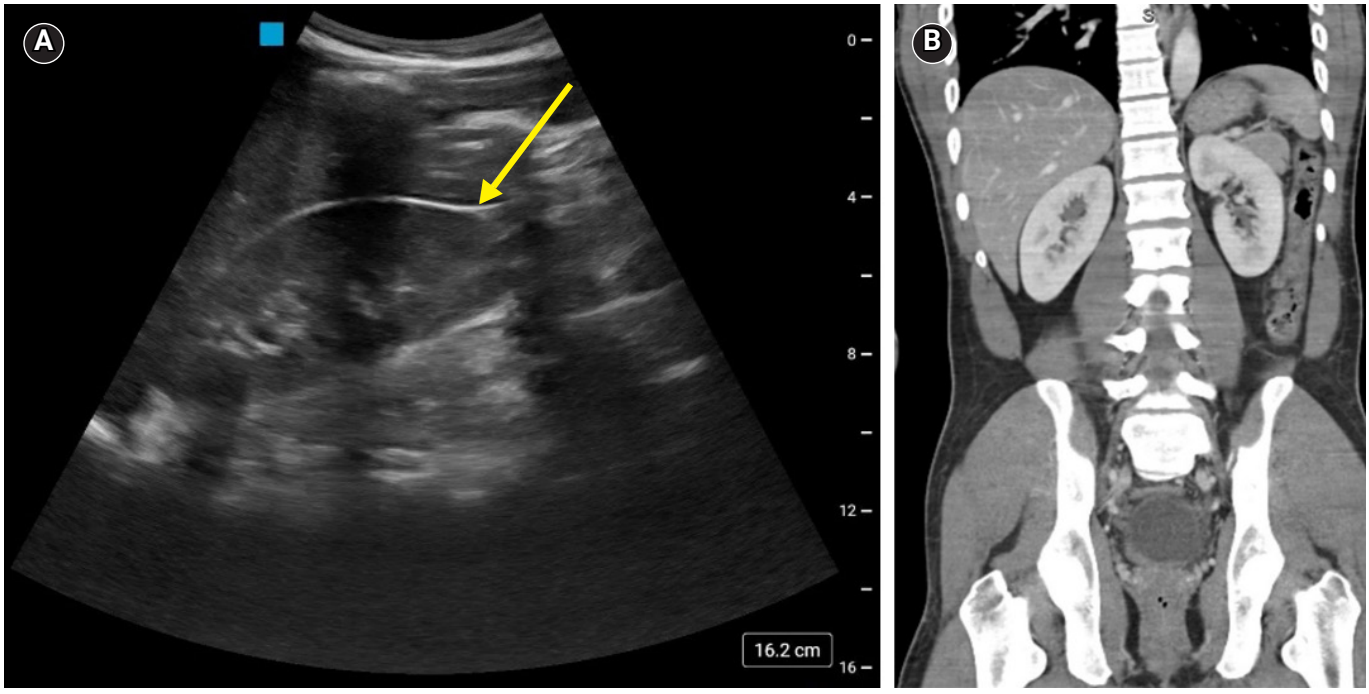


Fig. 1. Images of the patient. (A) An artifact creating a false anechoic stripe between the liver and kidney on the Focused Assessment with Sonography in Trauma (FAST) examination (arrow), caused by post-processing imaging enhancements. (B) Normal computed tomography image, demonstrating no intra-abdominal injury or free fluid.

adjustments can decrease the appearance of this artifact, there is a tradeoff in image quality. Without such changes, awareness of this artifact is critical to avoid unnecessary CT imaging or surgery.

ARTICLE INFORMATION

Ethics statement
Informed consent was waived due to the retrospective nature of this study and no personal identification is on the images provided.

Author contributions
Conceptualization: all authors; Investigation: all authors; Writing–original draft: LT, MG; Writing–review & editing: all authors. All authors read and approved the final manuscript.

Conflicts of interest
Michael Gottlieb currently receives grant funding through the US Centers for Disease Control and Prevention, Bill and Melinda Gates Foundation, Biomedical Advanced Research and Development Authority, and the Society of Directors of Research in Medical Education. He is also an editorial board member of this journal, but was not involved in the peer reviewer selection, evaluation, or decision process of this article. The authors have no other conflicts of interest to declare.

Funding
The authors received no financial support for this study.

Data availability
Data sharing is not applicable as no new data were created or analyzed in this study.

REFERENCES

1. Patel NY, Riherd JM. Focused Assessment with Sonography in Trauma: methods, accuracy, and indications. Surg Clin North Am 2011;91:195–207.
2. Lobo V, Hunter-Behrend M, Cullnan E, et al. Caudal edge of the liver in the right upper quadrant (RUQ) view is the most sensitive area for free fluid on the FAST exam. West J Emerg Med 2017;18:270–80.
3. Parker MA, Hicks BG, Kaili M, et al. The lipliner sign: potential cause of a false positive Focused Assessment with Sonography in Trauma (FAST) examination. J Emerg Med 2024;67:e553–9.