ORIGINAL ARTICLE

Performance of ChatGPT-4o in thoracic trauma: A comparative evaluation with guidelines

İsmail Dal,[1] Mehmet Yildirim[2]

[1]Department of Thoracic Surgery, Kastamonu University, Kastamonu-*Türkiye*

[2]Department of Thoracic Surgery, Dr. Siyami Ersek Thoracic and Cardiovascular Surgery Training and Research Hospital, Istanbul-*Türkiye*

ABSTRACT

BACKGROUND: This study aims to evaluate the performance of ChatGPT-4o in thoracic trauma management by comparing its responses to established clinical guidelines.

METHODS: Five major thoracic surgery guidelines were reviewed, including the Advanced Trauma Life Support (ATLS) Guidelines 2018, Eastern Association for the Surgery of Trauma (EAST) Guidelines 2020, Evaluation and management of traumatic pneumothorax: A Western Trauma Association critical decisions algorithm 2022, European Trauma Course (ETC) Guidelines 2016, and the National Institute for Health and Care Excellence (NICE) Guidelines for Trauma 2020. Fifty open-ended questions were developed based on these guidelines and submitted to ChatGPT-4o. Five thoracic surgery specialists evaluated the artificial intelligence (AI) responses using a 5-point Likert scale.

RESULTS: ChatGPT-4o achieved an average score of 4.76±0.57 on the 50-question evaluation. ChatGPT-4o excelled in questions derived from well-defined guidelines, demonstrating its ability to synthesize and apply guideline-based medical knowledge. Its performance aligns with previous studies in urological trauma and emergency medicine, which reported similar reliability. However, its reliance on pre-existing data limits its effectiveness in addressing highly nuanced or novel clinical scenarios. These findings underscore its potential as a complementary tool in guideline-driven medical contexts while emphasizing the need for clinical oversight in complex cases.

CONCLUSION: ChatGPT-4o performed strongly in thoracic trauma management questions, demonstrating minimal errors and high reliability. These results suggest it could serve as a valuable support tool for clinical decision-making, particularly in scenarios guided by established protocols. Further exploration into broader medical domains is warranted.

Keywords: Thoracic trauma; clinical decision support systems; artificial intelligence in medicine; guideline adherence; natural language processing.

INTRODUCTION

Thoracic trauma is a significant cause of morbidity and mortality worldwide, accounting for up to 25% of trauma-related deaths.[1,2] Prompt and accurate diagnosis, coupled with timely intervention, is critical for improving outcomes. Over the past two decades, various clinical guidelines have been developed to standardize the management of thoracic trauma.[3-7]

Artificial intelligence (AI) and natural language processing (NLP) tools have emerged as innovative aids in medical decision-making. AI models such as OpenAI's ChatGPT have shown promise in synthesizing and contextualizing large volumes of medical literature.[8,9] Previous studies have evaluated ChatGPT's performance in various medical specialties, including its role in urological trauma[10] and clinical oncology,[11] demonstrating its potential to support clinicians in decision-

Cite this article as: Dal İ, Yildirim M. Performance of ChatGPT-4o in thoracic trauma: A comparative evaluation with guidelines. Ulus Travma Acil Cerrahi Derg 2025;31:839-846.

Address for correspondence: İsmail Dal

Department of Thoracic Surgery, Kastamonu University, Kastamonu, Türkiye

E-mail: idal@kastamonu.edu.tr

Ulus Travma Acil Cerrahi Derg 2025;31(9):839-846 DOI: 10.14744/tjtes.2025.47087

Submitted: 01.01.2025 Revised: 23.07.2025 Accepted: 24.07.2025 Published: 05.09.2025

OPEN ACCESS This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

making. However, its performance in the domain of thoracic trauma remains underexplored.

Thoracic injuries encompass a wide spectrum, ranging from simple rib fractures to life-threatening conditions like tension pneumothorax and flail chest. The complexity of thoracic trauma management is compounded by its dependence on rapidly evolving evidence-based practices.[12] Guideline adherence has been shown to improve outcomes, but variations in practice persist due to factors such as lack of awareness, clinical inertia, and resource limitations.[13] This underscores the need for tools that can facilitate access to guideline-based information and enhance clinical decision-making.

Recent evaluations of AI models suggest their utility in addressing these gaps by providing clinicians with concise, evidence-based answers to complex clinical questions.[14,15] For instance, ChatGPT has been reported to perform well in translating and summarizing clinical guidelines into actionable recommendations.[16] Additionally, its ability to engage in interactive dialogue has the potential to simulate case-based discussions, which are crucial in trauma scenarios where rapid decisions are required.[17]

In this study, we sought to evaluate the performance of ChatGPT-4o in addressing thoracic trauma-related clinical questions derived from five widely recognized guidelines, including the Advanced Trauma Life Support (ATLS) Guidelines, the Eastern Association for the Surgery of Trauma (EAST), the European Trauma Course (ETC) Guidelines, the National Institute for Health and Care Excellence (NICE) Guidelines, and the Western Trauma Association Critical Decisions Algorithm.[3-7] By comparing its responses to expert opinions, this study aims to assess the model's capability to provide accurate, guideline-concordant answers and to explore its potential role in augmenting clinical practice in thoracic trauma management.

MATERIALS AND METHODS

Study Design

A descriptive cross-sectional study was designed to compare ChatGPT-4o's responses with guideline-based recommendations. The study involved generating clinical questions from established guidelines, obtaining answers from ChatGPT-4o, and evaluating these answers using a standardized scoring system reviewed by thoracic surgery experts.

Data Source and Question Development

Five well-recognized thoracic trauma management guidelines were selected as the basis for question development:

- Advanced Trauma Life Support (ATLS) Guidelines, 2018.[3]
- Eastern Association for the Surgery of Trauma (EAST) Practice Management Guidelines, 2020.[4]
- Evaluation and Management of Traumatic Pneumothorax: A Western Trauma Association Critical Decisions Algorithm, 2022.[5]
- European Trauma Course (ETC) Guidelines, 2016.[6]
- National Institute for Health and Care Excellence (NICE) Guidelines for Trauma, 2020.[7]

Fifty open-ended clinical questions addressing critical aspects of thoracic trauma management (e.g., pneumothorax, hemothorax, flail chest, rib fractures) were developed by two thoracic surgery specialists who are also authors of this study. Both had completed formal specialty training in thoracic surgery and had active clinical experience in the management of thoracic trauma. Their direct involvement in the design and objectives of the study ensured that each question was clinically relevant and explicitly linked to one or more of the selected guidelines, thereby maintaining both accuracy and comprehensiveness.

To illustrate the structure and clinical relevance of the questions and corresponding ChatGPT-4o responses, a representative question-answer pair is presented below.

Example Question:

Where is the appropriate surgical site for tube thoracostomy?

ChatGPT-4o Response:

Surgical Site:

Location: The fifth intercostal space, anterior to the midaxillary line (often referred to as the "safe triangle").

Landmarks: The tube should be inserted above the rib to avoid the neurovascular bundle that runs along the bottom of each rib. The site is chosen to allow drainage of air or fluid from the pleural space.

ChatGPT-4o Query Process

Each of the 50 questions was entered into ChatGPT-4o using a standardized format. The model was instructed to provide concise, evidence-based answers relevant to the question. No additional prompts or contextual clarifications were provided to simulate real-world clinical use.

All queries were conducted in December 2024 using the ChatGPT web interface with the default GPT-4o model available at that time (GPT-4o, May 2024 release). Each clinical question was submitted only once to simulate real-world conditions, particularly in time-sensitive clinical settings where multiple iterations are not feasible. No adjustments were made to model parameters such as temperature or randomness, as these settings are not user-configurable through the ChatGPT-4o web platform.

Evaluation of Responses

The responses were evaluated by five thoracic surgery spe-

Table 1. Evaluation of clinical consensus among experts

	Doctor 1	Doctor 2	Doctor 3	Doctor 4	Doctor 5	Mean	Standard Deviation
How many years have you been working as a thoracic surgery specialist?	1-10 Years	1-10 Years	1-10 Years	More than 20 years	1-10 Years		
1. Which pathologies should be identified and treated during the Primary Survey in thoracic trauma?	5	5	5	5	5	5	0.00
2. Which pathologies should be identified and treated during the Secondary Survey in thoracic trauma?	5	5	5	5	3	4.6	0.89
3. How should tension pneumothorax be diagnosed, and what is the treatment?	5	5	5	5	5	5	0.00
4. What clinical findings are observed in tension pneumothorax?	5	5	5	5	4	4.8	0.45
5. How are the respiratory mechanics disrupted in open pneumothorax?	5	5	5	5	4	4.8	0.45
6. What is the first intervention for open pneumothorax outside the hospital?	5	5	5	5	4	4.8	0.45
7. How do you clinically differentiate between tension pneumothorax and massive hemothorax?	5	5	5	5	5	5	0.00
8. How is the diagnosis of pericardial tamponade made following thoracic trauma?	5	5	5	5	5	5	0.00
9. How can tension pneumothorax and pericardial tamponade be differentiated through physical examination?	5	5	5	5	4	4.8	0.45
10. How should a patient be managed if they develop circulatory arrest and no pulse is palpable after thoracic trauma?	5	5	4	5	5	4.8	0.45
11. What should be done before air transport of a patient with pneumothorax?	5	5	4	5	5	4.8	0.45
12. What are the indications for thoracotomy in a patient with hemothorax following thoracic trauma?	5	5	4	5	5	4.8	0.45
13. How is blunt cardiac injury diagnosed, and what is the follow-up and treatment?	5	5	5	5	4	4.8	0.45
14. What are the chest X-ray findings in blunt aortic injury?	5	5	5	5	4	4.8	0.45
15. On which side is traumatic diaphragmatic rupture most common? How is it diagnosed, and what are the treatment principles?	5	5	5	5	5	5	0.00
16. What is the mechanism of blunt esophageal rupture, and how is it diagnosed? What are the treatment principles?	5	5	5	5	4	4.8	0.45
17. What is the significance of subcutaneous emphysema following thoracic trauma?	5	5	5	5	5	5	0.00

18. What complications can develop after rib fractures?	5	5	5	5	5	5	0.00
19. How do lower rib fractures (10th-12th) differ from other rib fractures??	5	5	5	5	5	5	0.00
20. How should a patient be managed if they have no palpable pulse but show signs of life after a penetrating thoracic injury?	5	5	5	5	5	5	0.00
21. How should a patient be managed if they have no palpable pulse and no signs of life after a penetrating thoracic injury?	5	5	5	5	4	4.8	0.45
22. Should emergency department thoracotomy be performed in a patient with extra-thoracic penetrating trauma who has no palpable pulse but shows signs of life?	5	5	3	5	5	4.6	0.89
23. Should emergency department thoracotomy be performed in a patient with extra-thoracic penetrating trauma who has no palpable pulse and no signs of life?	5	5	5	5	3	4.6	0.89
24. How should a patient be managed if they have no palpable pulse but show signs of life after blunt thoracic trauma?	5	5	5	5	4	4.8	0.45
25. How should a patient be managed if they have no palpable pulse and no signs of life after blunt thoracic trauma?	5	5	5	5	4	4.8	0.45
26. In pneumothorax diagnosis, should inspiratory or expiratory X-rays be preferred?	5	5	5	5	4	4.8	0.45
27. In a hemodynamically stable patient with traumatic pneumothorax, what criteria on X-ray and computed tomography (CT) should indicate the need for tube thora-costomy?	5	5	5	4	4	4.6	0.55
28. How is the differentiation made between small and large pneumothorax?	5	5	1	5	4	4	1.73
29. If a patient with pneumothorax has shortness of breath, should an intervention (needle aspiration, chest tube) be performed regardless of pneumothorax size?	5	5	5	5	4	4.8	0.45
30. How should a small pneumothorax without shortness of breath be managed?	5	5	5	5	5	5	0.00
31. Is needle aspiration associated with a shorter hospital stay compared to chest tubes in pneumothorax management?	5	4	5	5	2	4.2	1.30
32. What should be done if needle aspiration fails in pneumothorax management?	5	5	5	5	5	5	0.00
33. Is there a role for large bore (>20 F) chest tubes in pneumothorax management?	5	5	5	5	2	4.4	1.34
34. Should chest tubes in pneumothorax management be routinely connected to a suction system?	5	5	5	5	4	4.8	0.45

35. Under what circumstances can a patient with a history of pneumothorax engage in scuba diving?	5	5	5	5	4	4.8	0.45
36. In pneumothorax management, which is more successful: surgical or chemical pleurodesis? Which patients are more suitable for each option?	5	5	5	5	4	4.8	0.45
37. What is the most suitable agent for chemical pleurodesis?	5	5	5	5	4	4.8	0.45
38. How is pneumothorax managed during pregnancy?	5	5	5	5	4	4.8	0.45
39. Is prophylactic antibiotic use necessary during chest tube placement for traumatic pneumothorax?	5	5	5	5	2	4.4	1.34
40. How would performing chemical pleurodesis on a patient with cystic fibrosis and pneumothorax impact a potential lung transplant?	5	5	4	5	4	4.6	0.55
41. Which imaging modality has a higher diagnostic value for pneumothorax: ultrasound or X-ray?	5	4	5	5	3	4.4	0.89
42. What are the indications for intubation in flail chest?	5	5	4	5	4	4.6	0.55
43. What radiological findings are observed in pulmonary contusion on a chest X-ray?	5	5	4	5	4	4.6	0.55
44. What is the significance of a first rib fracture?	5	5	5	5	4	4.8	0.45
45. Where is the appropriate surgical site for tube thoracostomy?	5	5	5	5	4	4.8	0.45
46. In the emergency department, if pneumothorax is not detected on extended Focused Assessment with Sonography for Trauma (eFAST) in thoracic trauma, can it be excluded?	5	5	5	5	5	5	0.00
47. Only which patients should undergo chest decompression (needle, finger, chest tube) in the pre-hospital setting?	5	5	5	5	4	4.8	0.45
48. Which imaging studies should be ordered first for a severely unstable patient with respiratory distress following chest trauma?	5	5	5	5	4	4.8	0.45
49. Is it appropriate to choose a CT scan as the first imaging modality for thoracic trauma in patients under 16 years old?	5	4	5	5	4	4.6	0.55
50. What is the first choice of analgesic following major trauma?	5	5	5	5	4	4.8	0.45

cialists who participated voluntarily, following an open call distributed via a professional communication network among thoracic surgeons. All reviewers were certified specialists with clinical experience in thoracic trauma management. Their years of experience in thoracic surgery—ranging from 1 to over 20 years—are detailed in Table 1. A 5-point Likert scale was used for scoring, where:

1: Strongly Disagree – Incorrect or irrelevant response.

2: Disagree – Partially correct but missing key details.

3: Neutral – Adequate but lacking depth.

4: Agree – Comprehensive and correct.

5: Strongly Agree – Comprehensive, correct, and clinically insightful.

Reviewers assessed the accuracy, relevance, and adherence of the answers to guideline recommendations. Discrepancies in scoring were resolved through consensus.

Statistical Analysis

The mean and standard deviation of scores were calculated to determine ChatGPT-4o's overall performance. Statistical significance was defined as $p<0.05$. Data were analyzed using SPSS version 26.0 (IBM Corp., Armonk, NY, USA). The free-marginal multirater kappa test was used to calculate inter-rater reliability, with the tool available at "http://justusrandolph.net/kappa/" utilized for the analysis.[18]

Ethical Considerations

This study did not involve human or animal subjects and thus did not require formal ethical approval. All data used were derived from publicly available clinical guidelines. Reviewer participation was voluntary, and no personal or identifiable information was collected. According to the Declaration of Helsinki, ethical committee approval was not required for this study.

RESULTS

The mean score across all 50 questions was 4.76±0.57, indicating a high level of accuracy and adherence to guideline-based recommendations (Table 1). Median scores for all questions were consistently 5.0, reflecting a high level of agreement among reviewers regarding the model's performance. The evaluation produced a free-marginal kappa value of 0.58, with a 95% confidence interval of [0.51, 0.65], reflecting an intermediate to good level of agreement among the raters.

ChatGPT-4o performed best on questions requiring clear, guideline-based answers. For instance, it provided highly accurate and consistent responses on topics such as the diagnosis of tension pneumothorax, which are determined by standard protocols. While its performance on questions requiring more nuanced clinical judgment was also generally strong, slightly more variability was observed. Overall, there was robust agreement among the evaluating doctors, as reflected by consistently high scores.

Questions on emergency management steps, such as needle decompression for tension pneumothorax, received perfect scores (5.0±0.0), emphasizing the tool's utility in high-stakes clinical scenarios.

In contrast, questions requiring the identification of clinical findings or differential diagnosis displayed slightly varied scores, with a mean of 4.6±0.89. This variability suggests that while ChatGPT-4o provides accurate answers, its depth in clinical reasoning could be further refined.

DISCUSSION

The findings of this study demonstrate that ChatGPT-4o performs exceptionally well in answering clinical questions related to thoracic trauma, with a mean score of 4.76±0.57 on a 5-point Likert scale, indicating high accuracy and adherence to established guidelines.

Previous studies have explored the role of AI in trauma-related domains with promising results. For example, ChatGPT has demonstrated robust performance in synthesizing urological trauma guidelines,[10] and AI models have been applied in emergency medicine for prioritizing acute interventions.[19] These findings support the applicability of large language models in structured, protocol-driven trauma care. However, this study is one of the first to systematically evaluate the application of ChatGPT-4o in thoracic trauma management, addressing a critical gap in the literature.

The model's ability to provide accurate and guideline-concordant responses reflects its advanced natural language processing capabilities. Questions involving straightforward diagnostic and management protocols, such as tension pneumothorax or chest tube placement, received perfect scores (5.0±0.0), showcasing ChatGPT-4o's strength in synthesizing well-established evidence. These findings are consistent with other specialties, where AI tools excel in processing explicit and structured data.[20]

Additionally, ChatGPT-4o's interactive nature enables dynamic querying, allowing clinicians to refine or expand upon initial responses. This adaptability could be particularly valuable in resource-limited settings or for less experienced practitioners who might benefit from supplemental guidance.[21] The high performance observed in this study highlights ChatGPT-4o's utility as a decision-support tool in thoracic trauma management. By synthesizing complex guidelines into actionable insights, it can enhance clinical efficiency, particularly in high-pressure environments such as trauma centers.[22]

Moreover, ChatGPT-4o's application extends beyond individual clinicians. It could potentially be integrated into electronic health record (EHR) systems to provide real-time recommendations during patient care.[23] However, AI should not replace clinical expertise but instead act as an adjunct, supporting decision-making while allowing clinicians to exercise their judgment.

CONCLUSION

This study underscores the potential of ChatGPT-4o as a reliable, guideline-concordant tool for thoracic trauma manage-

ment. Its high performance, particularly in questions derived from explicit guidelines, demonstrates its promise in augmenting clinical decision-making. However, careful consideration of its limitations and integration with clinical expertise is essential to maximize its impact on patient care.

Ethics Committee Approval: This study did not involve human or animal subjects and thus did not require formal ethical approval.

Peer-review: Externally peer-reviewed.

Authorship Contributions: Concept: İ.D., M.Y.; Design: İ.D., M.Y.; Supervision: İ.D., M.Y.; Resource: İ.D., M.Y.; Materials: İ.D., M.Y.; Data collection and/or processing: İ.D., M.Y.; Analysis and/or interpretation: İ.D., M.Y.; Literature review: İ.D., M.Y.; Writing: İ.D., M.Y.; Critical review: İ.D., M.Y.

Conflict of Interest: None declared.

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

REFERENCES

1. Beshay M, Mertzlufft F, Kottkamp HW, Reymond M, Schmid RA, Branscheid D, et al. Analysis of risk factors in thoracic trauma patients with a comparison of a modern trauma centre: a mono-centre study. World J Emerg Surg 2020;15(1):45. [CrossRef]
2. Lundin A, Akram SK, Berg L, Göransson KE, Enocson A. Thoracic injuries in trauma patients: epidemiology and its influence on mortality. Scand J Trauma Resusc Emerg Med 2022;30(1):69. [CrossRef]
3. Advanced Trauma Life Support. Advanced Trauma Life Support (ATLS®) Student Course Manual. 10th ed. Chicago: American College of Surgeons; 2018.
4. Eastern Association for the Surgery of Trauma. EAST Practice Management Guidelines for Trauma. 2020.
5. de Moya M, Brasel KJ, Brown CVR, Hartwell JL, Inaba K, Ley EJ, et al. Evaluation and management of traumatic pneumothorax: A Western Trauma Association critical decisions algorithm. J Trauma Acute Care Surg 2022;92(1):103-7.[CrossRef]
6. European Trauma Course. European Trauma Course Manual. 4th ed. London: European Trauma Course Organisation; 2016.
7. National Institute for Health and Care Excellence. NICE guidelines for trauma. London: NICE; 2020.
8. 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. [CrossRef]
9. Tian S, Jin Q, Yeganova L, Lai PT, Zhu Q, Chen X, et al. Opportunities and challenges for ChatGPT and large language models in biomedicine and health. Brief Bioinform 2023;25(1):bbad493. [CrossRef]
10. Li J, Yi X, Han Z, Liao D, Zhang T, Liu Let al. The theranostic performance of Chat-GPT against urological trauma. Int J Surg 2024;110(7):4485-7. [CrossRef]
11. Kann BH, Hosny A, Aerts HJWL. Artificial intelligence for clinical oncology. Cancer Cell 2021;39(7):916-27. [CrossRef]
12. Fitzpatrick ER. Evidence-Based Pearls: Chest Trauma. Crit Care Nurs Clin North Am 2023 Jun;35(2):129-44. [CrossRef]
13. Vasse M, Leone M, Boyer L, Michelet P, Goudard Y, Cardinale M, et al. Impact of the implementation of a trauma system on compliance with evidence-based clinical management guidelines in penetrating thoracic trauma. Eur J Trauma Emerg Surg 2023;49(1):351-60. [CrossRef]
14. Verma K, Moore M, Wottrich S, et al. Emulating Human Cognitive Processes for Expert-Level Medical Question-Answering with Large Language Models. arXiv [preprint] arXiv:2310.11266.2023.
15. Vaid A, Lampert J, Lee J, et al. Generative Large Language Models are Autonomous Practitioners of Evidence-Based Medicine. arXiv [preprint] 2024:2401.02851.
16. Wang Y, Visweswaran S, Kapoor S, et al. ChatGPT, Enhanced with Clinical Practice Guidelines, is a Superior Decision Support Tool. medRxiv [preprint] 2023.
17. Jananee V, Nivedhitha GK, Pujasree K. AI Powered Chat Assistant for Trauma Detection from Text and Voice Conversations with a Direct Doctor Connection. In: Vimal V, Perikos I, Mukherjee A, Piuri V, editors. Multi-Strategy Learning Environment. Singapore: Springer Nature Singapore; 2024. p.118. [CrossRef]
18. Randolph J. Free-Marginal Multirater Kappa (multirater κfree): An Alternative to Fleiss Fixed-Marginal Multirater Kappa. Adv Data Anal and Classif 2010;4.
19. Piliuk K, Tomforde S. Artificial intelligence in emergency medicine. A systematic literature review. Int J Med Inform 2023;180:105274. [CrossRef]
20. Liu PR, Lu L, Zhang JY, Huo TT, Liu SX, Ye ZW. Application of Artificial Intelligence in Medicine: An Overview. Curr Med Sci 2021;41(6):1105-15. [CrossRef]
21. Elhaddad M, Hamam S. AI-Driven Clinical Decision Support Systems: An Ongoing Pursuit of Potential. Cureus 2024;16(4):e57728. [CrossRef]
22. Aziz HA, Lunde J, Barraco R, et al. Evidence-based review of trauma center care and routine palliative care processes for geriatric trauma patients; A collaboration from the American Association for the Surgery of Trauma Patient Assessment Committee, the American Association for the Surgery of Trauma Geriatric Trauma Committee, and the Eastern Association for the Surgery of Trauma Guidelines Committee. J Trauma Acute Care Surg 2019;86(4):737-43. [CrossRef]
23. Schwieger A, Angst K, de Bardeci M, Burrer A, Cathomas F, Ferrea S, et al. Large language models can support generation of standardized discharge summaries - A retrospective study utilizing ChatGPT-4 and electronic health records. Int J Med Inform 2024;192:105654. [CrossRef]

ORİJİNAL ÇALIŞMA - *ÖZ*

Torasik travmada ChatGPT-4o performansı: Kılavuzlara dayalı karşılaştırmalı değerlendirme

AMAÇ: Bu çalışma, Chat GPT-4o'nun torakal travma yönetimindeki performansını, yanıtlarını belirli klinik kılavuzlarla karşılaştırarak değerlendirmeyi amaçlamaktadır.

GEREÇ VE YÖNTEM: Beş ana torakal cerrahi kılavuz incelendi; bunlar arasında 2018 yılına ait Advanced Trauma Life Support (ATLS) Kılavuzları, 2020 yılına ait Eastern Association for the Surgery of Trauma (EAST) Kılavuzları, 2022 yılına ait Evaluation and Management of Traumatic Pneumothorax: A Western Trauma Association Critical Decisions Algorithm, 2016 yılına ait European Trauma Course (ETC) Kılavuzları ve 2020 yılına ait NICE Trauma Kılavuzları yer almaktadır. Bu kılavuzlara dayalı olarak 50 açık uçlu soru geliştirilmiş ve Chat GPT-4o'ya sunulmuştur. Beş torakal cerrahi uzmanı, yapay zekânın yanıtlarını 5 dereceli Likert ölçeği ile değerlendirmiştir.

BULGULAR: Chat GPT-4o, 50 soruluk değerlendirmede ortalama 4.76±0.57 puan almıştır. Chat GPT-4o, iyi tanımlanmış kılavuzlardan türetilen sorularda başarılı olmuş ve kılavuzlara dayalı tıbbi bilgiyi sentezleme ve uygulama yeteneğini göstermiştir. Performansı, ürolojik travma ve acil tıp alanındaki önceki çalışmalarla uyumludur ve benzer güvenilirlik bildirilmiştir. Ancak, mevcut verilere dayalı çalışması, oldukça ayrıntılı veya yeni klinik senaryoları ele almadaki etkinliğini sınırlamaktadır. Bu bulgular, kılavuz odaklı tıbbi bağlamlarda, özellikle karmaşık vakalarda klinik denetim gerekliliğine vurgu yaparak, onun tamamlayıcı bir araç olarak potansiyelini vurgulamaktadır.

SONUÇ: Chat GPT-4o, torakal travma yönetimi sorularında güçlü bir performans sergilemiş, hataları minimumda tutmuş ve yüksek güvenilirlik göstermiştir. Bu sonuçlar, özellikle belirlenmiş protokollerle yönlendirilen senaryolarda, klinik karar verme sürecinde değerli bir destek aracı olabileceğini göstermektedir. Daha geniş tıbbi alanlarda daha fazla araştırma yapılması gereklidir.

Anahtar sözcükler: Doğal dil işleme; klinik karar destek sistemleri; kılavuzlara uyum; tıpta yapay zeka; torakal travma.

Ulus Travma Acil Cerrahi Derg 2025;31(9):839-846 DOI: 10.14744/tjtes.2025.47087