Turkish Journal of Clinics and Laboratory

To cite this article: Gokcinar D, Akan B, Kursun O, Gokcinar AH, Akan B. ChatGPT 4 may inappropriately answer questions about managing critically ill patients. Turk J Clin Lab 2025; 4: 688-692.

Letter to the Editor

ChatGPT 4 may inappropriately answer questions about managing critically ill patients

ChatGPT 4 kritik hastaları yönetmekle ilgili soruları uygunsuz bir şekilde yanıtlayabilir

Derya Gokcinar*[1], Belgin Akan[1], Oguzhan Kursun[2], Abdullah Halit Gokcinar[3], Baris Akan[4]

[1]Department of Critical Care, University of Health Sciences, Ankara Bilkent City Hospital, Ankara, Turkey
[2]Department of Neurology, University of Health Sciences, Ankara Bilkent City Hospital, Ankara, Turkey
[3]Gap Incorporated, San Francisco, Illinois, United States of America
[4]TOBB University of Economics and Technology, Ankara, Türkiye

Abstract

Intensive care patients have extremely complex pathologies, making their treatment management difficult. In this study, we aim to evaluate ChatGPT 4, a large language model, regarding its knowledge of critically ill patient management. Scenarios involving mechanical ventilation were created by an intensivist, anesthesiologist, and neurologist working in intensive care units, including head trauma, pulmonary embolism, myocardial infarction, chronic obstructive pulmonary disease, chronic kidney disease, and infective endocarditis, septic shock, and status epilepticus. Questions about patient management related to these scenarios were proposed to ChatGPT 4. Although ChatGPT 4 answered most of the questions correctly, it still requires medical input.

Keywords: large language models, ChatGPT 4, intensive care units, critical care, artificial intelligence

Öz

Yoğun bakım hastaları son derece karmaşık patolojilere sahiptir ve bu da tedavi yönetimlerini zorlaştırır. Bu çalışmada, büyük dilli bir model olan ChatGPT 4'ü kritik hasta yönetimi konusundaki bilgi birikimi açısından değerlendirmeyi amaçladık. Yoğun bakım ünitelerinde çalışan bir yoğun bakım uzmanı, anestezist ve nörolog tarafından kafa travması, pulmoner emboli, miyokard enfarktüsü, kronik obstrüktif akciğer hastalığı, kronik böbrek hastalığı ve enfektif endokardit, septik şok ve status epileptikus dahil olmak üzere mekanik ventilasyon içeren senaryolar oluşturuldu. Bu senaryolarla ilgili hasta yönetimiyle ilgili sorular ChatGPT 4'e önerildi. ChatGPT 4 soruların çoğunu doğru yanıtlasa da, yine de tıbbi girdi gerektiriyor.

Anahtar kelimeler: Büyük dil modelleri, ChatGPT 4, yoğun bakım üniteleri, kritik bakım, yapay zeka

Corresponding Author*: Derya Gokcinar. Ankara Bilkent City Hospital, Department of Critical Care, Street 1604, Number 9, Postal code 06800, Çankaya, Ankara, Türkiye.
Email: deryaozgurgokcinar@gmail.com Phone: +905052283639
ORCID ID: 0000-0002-0870-2168
Doi: 10.18663/tjcl.1762561
Received: 11.08.2025 accepted: 13.10.2025

Öz

Amaç: Postherpetik nevralji (PHN), herpes zoster enfeksiyonunun yaşam kalitesini önemli ölçüde etkileyen en yaygın komplikasyonudur. Bu bağlamda, bu çalışmanın amacı serum 25-hidroksivitamin D [25(OH)D] düzeyleri ile PHN'nin şiddeti, süresi ve nöropatik semptomları arasındaki ilişkiyi araştırmaktır.

Gereç ve Yöntemler: Bu retrospektif gözlemsel kohort çalışmasının örneklemi, Ocak 2020 ile Haziran 2024 arasında PHN tanısı ile takip edilen 71 hastadan oluşmaktadır. Hastalar serum 25(OH)D düzeylerine göre D vitamini eksikliği (≤20 ng/mL, n=24), D vitamini yetersizliği (21-29 ng/mL, n=26) ve D vitamini yeterliliği (≥30 ng/mL, n=21) olan hastalar olarak üç gruba ayrılmıştır. Hastaların nöropatik ağrıları nöropatik ağrı 4 (DN4) anketi ve Numerik derecelendirme ölçeği (NRS) kullanılarak değerlendirilmiştir. Her hasta için C-reaktif protein (CRP) düzeyi, eritrosit sedimentasyon hızı (ESR) ve lökosit sayısı gibi inflamatuar parametreler kaydedilmiş ve serum 25(OH)D düzeylerinin tanısal performansı alıcı işletim karakteristiği (ROC) eğrisi analizi ile değerlendirilmiştir.

Bulgular: D vitamini eksikliği olan hastaların ortalama DN4 ve NRS skorları ile ağrı süreleri diğer hastalara göre anlamlı derecede yüksekti (5,83±1,76'ya karşı 1,71±0,96, p<0,001; 8,92±1,59'a karşı 6,29±2,43, p=0,001; 93,54±41,23'e karşı 31,86±35,80 gün, p<0,001; sırasıyla). Serum 25(OH)D düzeyi, DN4 skoru (r=-0,758; p<0,001), ağrı süresi (r=-0,640; p<0,001) ve NRS skoru (r=-0,497; p<0,001) ile negatif yönde güçlü korelasyon gösterdi. İğnelenme hissi ve fırçayla uyarılan ağrı, D vitamini eksikliği olan hastalarda diğer hastalara göre anlamlı derecede daha yaygındı (p<0,001). D vitamini eksikliği olan hastaların ortalama CRP ve ESR değerleri diğer hastalara göre anlamlı derecede yüksekti (11,99±14,56'ya karşı 6,99±12,38 mg/L, p=0,046 ve 25,71±16,58'e karşı 13,75±7,01 mm/saat, p=0,007; sırasıyla). ROC analizleri, DN4 skoru ≥4 olan hastaları ve iğnelenme hissini tahmin etmek için optimal serum 25(OH)D kesme düzeylerinin sırasıyla 21,15 ng/mL (eğri altında kalan alan [AUC]=0,862) ve 25,85 ng/mL (AUC=0,930) olduğunu ortaya koydu.

Sonuç: Düşük serum D vitamini düzeyleri PHN hastalarında artmış nöropatik ağrı yoğunluğu, uzamış ağrı süresi ve artmış inflamatuar yanıt ile ilişkilendirilmiştir. Sonuç olarak, D vitamini düzeyi PHN'nin klinik seyrini tahmin etmek ve uygun tedavi stratejilerini belirlemek için değerli bir biyobelirteç olabilir.

Anahtar kelimeler: postherpetik nevralji; d vitamini eksikliği; dn4 skoru; nöropatik ağrı; herpes zoster

Dear Editor,

Advanced life support and complex treatments are applied to patients admitted to intensive care units. Treatments and interventions that are unavailable from other services are administered in these units under close supervision. Here, patients are monitored more frequently and in detail. Mechanical ventilation and extracorporeal organ support treatments are performed under the supervision of trained personnel. Medical treatments can be complex and should be performed swiftly alongside important intervention and treatment decisions. Treatments that adjust according to frequently changing patient vitals are required. Physicians working in intensive care units undergo years of special training. Knowledge and skills are important because an individualist approach is applied to patients. Today, developments in machine learning, artificial intelligence, and robotics are promising benefits for medicine. The support of artificial intelligence in the treatment of intensive care patients could provide better quality service and results. Large language models such as ChatGPT 4 are applied in various fields. ChatGPT 4 was asked how it should be used in the field of intensive care; its answer included how it can aid with information and clinical decision-making [1].

We considered whether ChatGPT 4 has sufficient knowledge of intensive care medicine to help with treatment options. In this study, we aimed to ask ChatGPT 4 questions about the treatments of different patient profiles admitted to intensive care.

Questions about scenarios involving critically ill patients were presented to ChatGPT 4. The patient scenarios were entirely fictional and did not contain real individuals' information. These scenarios included common reasons for ICU admission, such as mechanical ventilation, head trauma, pulmonary embolism, myocardial infarction, chronic obstructive pulmonary disease, chronic kidney disease, infective endocarditis, septic shock, and status epilepticus. These scenarios were created by an intensivist, an anesthesiologist, and a neurologist with more than 10 years of experience working in intensive care units. Questions were formulated about the treatment management of patients by these physicians to ask ChatGPT 4. These three experts analyzed the responses from ChatGPT 4. This study does not require copyrighting for ChatGPT 4 outputs. According to OpenAI's Content Policy and Terms of Use, ChatGPT users own all outputs they create with large language models, including the text and images.

We obtained ChatGPT 4's responses to our designed questions. ChatGPT's answers to our questions are available at: https://github.com/barisaakan/llm-patients [2].

Table 1. Scenarios, questions, and responses are listed by case type.

Case number and type	Patient scenarios and questions	Evaluation of ChatGPT 4's answers
Case Example 1 Mechanical ventilation	A 57-year-old male patient, 175 cm tall, diagnosed with pneumonia, is on mechanical ventilation. His lung compliance is measured at 27 ml/cmH2O, his driving pressure is 19 cmH2O, and his stress index is above 1. The airway peak pressure is 31 cmH2O, PEEP is set at 5 cmH2O, and tidal volume is 375 ml. The patient is in a volume-controlled ventilation mode. Question A) How would you adjust the mechanical ventilator settings? Question B) In this patient, could the high flow in a pressure-controlled or PRVC mode have harmful effects? Question C) Could volume-controlled ventilation cause a pendelluft effect in this patient?	ChatGPT 4 suggested the following for the first case example: "Use a decelerating inspiratory flow pattern (most lung-protective)." ChatGPT 4's suggestion of a decelerating inspiratory flow pattern without knowledge of both the patient and our mechanical ventilator is incorrect. It makes suggestions without knowledge of the brand, modes, and flow patterns that our specific mechanical ventilator can apply. It has been shown that decelerating inspiratory flow patterns cause higher peak flow compared to sinusoidal, square, and trunk decelerating waveforms in ARDS patients [3,4].
Case Example 2 Head trauma	Question A) What would your treatment recommendation be for a patient admitted to intensive care due to head trauma and connected to mechanical ventilation? Question B) If a patient with head trauma has urine output exceeding 200 ml per hour, what would you consider? Question C) How should the treatment be managed?	ChatGPT 4 made accurate suggestions for case example 2.
Case Example 3 Pulmonary embolism	A 61-year-old male patient, 135 kg, 184 cm tall, was brought to the emergency department due to the sudden onset of respiratory distress, hypotension, and syncope about an hour ago. A thoracic CT angiography detected massive pulmonary embolism. Blood gas analysis: pH 7.42, PO2 48.5 mmHg, PCO2 40.9 mmHg, SO2 82.7%, BEB -0.03. Question A) What are your treatment recommendations? Question B) What are the contraindications for thrombotic therapy?	ChatGPT 4 made accurate suggestions for case example 3.
Case Example 4 Myocardial infarction	A 40-year-old male patient, who has smoked for 25 years, presented to the emergency department with severe, pressure-like chest pain and numbness in his left arm. ECG showed ST-segment elevations in anterior leads. He was admitted to intensive care. Blood pressure: 110/67 mmHg; heart rate: 92 bpm; respiratory rate: 18/min; peripheral oxygen saturation: 98%; body temperature: 36.5°C. Question A) What is your treatment recommendation? Question B) What are the risks of percutaneous coronary interventions?	ChatGPT 4 made accurate suggestions for case example 4.
Case Example 5 Chronic obstructive pulmonary disease	A 60-year-old female patient with chronic obstructive pulmonary disease, essential hypertension, atherosclerotic heart disease, and heart failure was admitted to the intensive care unit due to respiratory distress. Thoracic CT showed bilateral pleural effusion with a 2 cm thickness, fine fibrotic changes in the lung parenchyma, and an enlarged heart size. Echocardiography showed a pulmonary trunk diameter of 40 mm, an ejection fraction of 30%, a left atrium size of 4.4 cm x 6.7 cm, and grade 2 mitral regurgitation. The apex and lateral wall of the left ventricle were hypokinetic. Question: What are your treatment recommendations for this patient?	ChatGPT 4 made accurate suggestions for case example 5.

Case Example 6 Chronic kidney disease and infective endocarditis	A 61-year-old male patient, 87 kg, 174 cm tall, with chronic kidney failure undergoing scheduled dialysis three times a week, was admitted to the intensive care unit due to a purulent infected arteriovenous fistula in his left arm, a fever of 39°C, fatigue, and respiratory distress. Abdominal CT showed a 3 cm x 2 cm infarct in the spleen. Echocardiography revealed a 2.5 cm x 1.9 cm infective endocarditis lesion on the mitral valve. The left atrium was measured at 4.9 cm, and grade 3 mitral regurgitation was detected. Blood cultures isolated Streptococcus pneumoniae in two separate samples. Hemodialysis was performed via a femoral venous catheter. Question A) What are your treatment recommendations? Question B) How would you adjust antibiotic doses according to dialysis? Question C) Would you recommend surgery for this patient?	In general, Chat GPT made treatment recommendations in accordance with the guidelines for case example 6, who had infective endocarditis [5]. However, ChatGPT 4's suggestion that "Anticoagulation should be avoided due to embolization risk" is not always valid. Personalized strategies are recommended for anticoagulation treatment in the management of infective endocarditis [6]. Perhaps the method of asking ChatGPT 4 questions by only entering some of the patient's data is ineffective. Physicians constantly monitor the patient's history, complaints, physical health, laboratory, and imaging results on the monitor. If we offer these opportunities to artificial intelligence and develop treatment algorithms, more successful disease management results could be achieved.
Case Example 7 Septic shock	A 72-year-old female patient with congestive heart failure, chronic kidney disease, chronic obstructive pulmonary disease, hypothyroidism, hypertension, and type 2 diabetes mellitus, measuring 155 cm in height and weighing 117 kg, was admitted to the intensive care unit due to septic shock. She was receiving norepinephrine infusion at 0.5 mcg/kg/min. She had widespread erythematous and purulent discharge lesions on her scalp, trunk, face, and legs. Cervical, axillary, and inguinal lymph nodes were enlarged and palpable. Thoracic CT showed increased reticulonodular activity. A punch biopsy from skin lesions revealed Mycosis fungoides with significant cell changes. Cultures from skin wounds showed the growth of Pseudomonas aeruginosa, which was sensitive to amikacin and meropenem. Blood test results: - White blood cell count: 19.3 x10^9/L, neutrophils: 81%; - Blood gas: PCO2: 56 mmHg, PO2: 54.4 mmHg, SO2: 65%, BEB: -5.0, HCO3: 23 mmol/L, pH: 7.22; - CRP: 231, Procalcitonin: 0.50 mcg/L, ESR: 67 mm/hr. Question A) How would you treat this patient? Question B) Later, a blood culture showed Klebsiella pneumoniae, which was resistant to all antibiotics except ceftazidime-avibactam. Urine output dropped below 35 ml/hour, hypotension worsened, and norepinephrine infusion increased to 1 mcg/kg/min. What are your treatment recommendations?	For case example 7, we found ChatGPT 4's "multidisciplinary ICU management" suggestion appropriate. However, Chat GPT itself predicted "potential multiorgan failure" even though we did not ask about the patient's prognosis. It stated that aggressive intervention is required and recommended, considering palliative care if refractory shock develops. As such, we interpreted the "very poor prognosis" prediction as interesting.
Case Example 8 Status epilepticus	A 19-year-old female patient with Lafora disease, receiving 1000 mg/day sodium valproate, 50 mg/day topiramate, and 1500 mg/day levetiracetam, was brought to the emergency department due to generalized seizures. She was administered 10 mg IV diazepam, intubated, and connected to mechanical ventilation before being admitted to the intensive care unit. The patient continued to exhibit widespread myoclonic jerks. Question: What is your treatment recommendation?	For case example 8, ChatGPT 4 recommended stopping sodium valproate and levetiracetam treatment. However, sodium valproate and levetiracetam are widely used in the treatment of Lafora disease and myoclonus. According to newly published patient results, valproic acid is beneficial in such cases and its use is recommended [7,8]. Sodium valproate is not used in mitochondrial diseases because it increases mitochondrial dysfunction. We think ChatGPT 4 may be confusing Lafora disease with mitochondrial disorders here. In fact, levetiracetam is not used as the main antiepileptic as it already has a similar structure to piracetam. High doses of pirecetam were used in the past, but additional treatment with levetirecetam is now recommended.

Table 1 shows the patient scenarios presented to ChatGPT 4 and its responses. We identified the correct answers to ChatGPT 4 questions 2, 3, 4, and 5. However, some answers were incomplete, false, or incorrect for questions 1, 6, 7, and 8.

ChatGPT 4 mostly gave appropriate answers to questions on case examples. It provided information about each case and suggested treatment methods. Here, ChatGPT 4 presented its findings as if they were completely evidence-based information. Treatment recommendations were presented as necessary despite a small number of studies with low evidence and a limited number of subjects. Sometimes, information was accumulated like a book; however, it could not personalize this information to the patient. It did not ask us important questions that could reveal more information about a patient's particular condition. Overall, we felt like we asked Mr. Spock from the Star Trek movie for a consultation on patients.

Meo et al. asked multiple-choice questions about medical sciences to ChatGPT 3.5, which answered most of the questions correctly with a score of 72%. They suggested that Chat GPT 3.5 can support the education of medical students [9]. However, could artificial intelligence also be useful in clinic settings? What solutions could artificial intelligence propose for the instantaneous changes seen in patients? In the present study, ChatGPT 4 did not answer some questions correctly. As such, we are interested in newly developed artificial intelligence models. Future studies should be conducted on these issues with the hope that humanoid robots can help in the treatment and care of patients.

In a study reported by Wang et al., ChatGPT 4 achieved a 73.5% success rate on a 600-question exam presented to critical care physicians in China. However, they expressed concern that ChatGPT 4 might use misleading information when in patient management. They recommended that the information provided by ChatGPT 4 be supervised by specialist physicians [10].

Artificial intelligence is currently at a level where it can answer questions with its own algorithm for the decision-making process. In the future, artificial intelligence could support the management of autonomous intensive care processes. Intensive care treatment and support systems are complex. The question of how to apply rapidly increasing intensive care knowledge and developments to decision-making processes is a significant issue that needs resolving. The training of intensive care physicians, nurses, and other staff is an extensive and laborious process.

In this study, ChatGPT 4 answered most of the questions correctly. With the advancement of machine learning in the medical field, intensive care management provided by AI has the potential to achieve beneficial outcomes in the future.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

This research received no external funding.

Ethics approval

Ethics approval is not necessary for this paper.

Consent

The patient scenarios described in this study are entirely fictitious. This study does not require ethics committee approval because human and animal data are not used.

Authors' contributions

All authors contributed to the study conception and design. The scenarios were created by Derya Gokcinar, Belgin Akan and Oguzhan Kursun. The questions were asked to ChatGPT 4 by Baris Akan. Derya Gokcinar, Belgin Akan and Oguzhan Kursun analyzed the responses from ChatGPT 4. The first draft of the manuscript was written by Derya Gokcinar and Abdullah Halit Gokcinar, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

1. Lu Y, Wu H, Qi S, Cheng K. Artificial Intelligence in Intensive Care Medicine: Toward a ChatGPT/GPT-4 Way? Ann Biomed Eng 2023; 51: 1898-903.
2. https://github.com/barisaakan/llm-patients
3. Monjezi M, Jamaati H. The effects of pressure- versus volume-controlled ventilation on ventilator work of breathing. Biomed Eng Online 2020; 19: 72.
4. Chiumello D, Meli A, Pozzi T, Lucenteforte M, Simili P, Sterchele E, Coppola S. Different Inspiratory Flow Waveform during Volume-Controlled Ventilation in ARDS Patients. J Clin Med 2021; 10: 4756.
5. McDonald EG, Aggrey G, Aslan AT, Casias M, Cortes-Penfield N, Dong MQD et al. Guidelines for Diagnosis and Management of Infective Endocarditis in Adults: A WikiGuidelines Group Consensus Statement. JAMA Netw Open 2023; 6: e2326366.
6. Zhu X, Wang Z, Ferrari MW, Ferrari-Kuehne K, Hsi et al. Management of anticoagulation in patients with infective endocarditis. Thromb Res 2023; 229: 15-25.
7. Ibrahim F, Murr NI. Lafora Disease 2024. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025.
8. Parihar R, Ganesh S. Lafora progressive myoclonus epilepsy: Disease mechanism and therapeutic attempts. J Biosci 2024; 49: 22.
9. Meo SA, Al-Masri AA, Alotaibi M, Meo MZS, Meo MOS. ChatGPT Knowledge Evaluation in Basic and Clinical Medical Sciences: Multiple Choice Question Examination-Based Performance. Healthcare (Basel) 2023; 11: 2046.
10. Wang X, Tang J, Feng Y, Tang C, Wang X. Can ChatGPT-4 perform as a competent physician based on the Chinese critical care examination? J Crit Care 2025; 86: 155010.