Original Article | Journal of Orthopaedic Case Reports 2026 June:16(6):Page 220-226

# Clinical Implications of ChatGPT-assisted Multimodal Pre-operative Assessment in Elderly Pertrochanteric Fracture Patients: An Exploratory Study

Mitsuaki Noda[1], Shunsuke Takahara[2], Shinya Hayashi[3], Atsuyuki Inui[3], Keisuke Oe[3], Takehiko Matsushita[4]

**Learning Point of the Article:**

ChatGPT-assisted multimodal integration of pre-operative data demonstrated reproducible separation between survivors and non-survivors, suggesting a potential framework for clinical pattern recognition rather than prediction.

## Abstract

**Introduction:** Given the high mortality associated with femoral trochanteric fractures, reliable pre-operative estimation of post-operative death risk is essential. Commonly used scoring systems demonstrate only moderate predictive ability, largely because they rely on limited and sometimes outdated clinical parameters. ChatGPT (Generative Pre-Training Transformer) may be capable of synthesizing diverse pre-operative clinical information within a single multimodal analytical framework. Therefore, this study aimed to (i) Explore whether ChatGPT-based multimodal integration of pre-operative data can generate clinically interpretable patterns and (ii) Assess reproducibility of generated outputs across two repeated evaluations.

**Materials and Methods:** Patients with pertrochanteric fractures were retrospectively reviewed. Demographic variables and image sets, including laboratory data, cardiac reports, and medication records, were uploaded to ChatGPT (GPT-4o and GPT-5). The same standardized prompt was used to generate a 2-month post-operative mortality risk (%) after multimodal integration of demographic and clinical image-based information, and each model was tested twice. Patients were classified as Alive or Dead based on 2-month post-operative status. Generated output values (%) were then examined for between-group separation. Agreement between repeated estimates was assessed using Bland-Altman analysis.

**Results:** A total of 134 patients were included. The Alive group comprised 129 patients (106 females; mean age, 87 years), and the Death group comprised 5 patients (3 females; mean age 90 years). ChatGPT-4o showed no significant difference between groups. GPT-5 generated higher output values in the Death group across both repeated evaluations, indicating separation between groups after multimodal data integration. Agreement analysis showed a mean difference (bias) of 1.28% (95% confidence interval [CI], 0.09–2.47) for GPT-4o and 0.18% (95% CI, −0.41–0.76) for GPT-5.

**Conclusions:** ChatGPT, particularly GPT-5, demonstrated separation of generated output values between survivors and non-survivors, after integrating diverse pre-operative data. However, this should not be interpreted as evidence of predictive accuracy, primarily due to the severe imbalance in group size. Rather, these findings suggest potential clinical relevance of AI-assisted multimodal assessment as a pathway for addressing complex medical questions in daily practice.

**Keywords:** Pertrochanteric fracture, pre-operative data, multimodal integration, ChatGPT, optical character recognition.

Access this article online

Website: www.jocr.co.in

DOI: https://doi.org/10.13107/jocr.2026.v16.i06.7428

*Author's Photo Gallery*


Dr. Mitsuaki Noda


Dr. Shunsuke Takahara


Dr. Shinya Hayashi


Dr. Atsuyuki Inui


Dr. Keisuke Oe


Dr. Takehiko Matsushita

[1]Department of Orthopedics, Himeji Central Hospital, Himeji, Japan,
[2]Department of Orthopedics, Hyogo Prefectural Kakogawa Hospital, Kakogawa, Japan,
[3]Department of Orthopedics, Kobe University Graduate School of Medicine, Kobe, Japan,
[4]Department of Orthopedics, Meiwa Hospital, Nishinomiya, Japan.

**Address of Correspondence:**
Dr. Mitsuaki Noda,
Department of Orthopedics, Himeji Central Hospital, Himeji, Japan.
**E-mail:** m-noda@muf.biglobe.ne.jp

Submitted: 05/03/2026; Review: 25/04/2026; Accepted: May 2026; Published: June 2026

DOI: https://doi.org/10.13107/jocr.2026.v16.i06.7428

## Introduction

Hip fractures represent a major health burden, reducing life expectancy in older adults by approximately 25% compared with age-matched controls [1]. Early post-operative mortality is a key outcome, with reported 30-day mortality rates ranging from 6.0% to 12.0% [2,3], and up to one-third of patients dying within the 1st post-operative year [3]. Surgery is usually justified because it relieves pain and facilitates early mobilization, even in medically fragile individuals [2]. However, given the high mortality risk, non-operative management may be a reasonable alternative in selected frail elderly patients when the likelihood of post-operative death appears substantial [2,4]. In addition, some local hospitals may recommend transferring to higher-level facilities when serious post-operative complications are anticipated. However, commonly used mortality prediction tools such as the Nottingham Hip Fracture Score, ASA Physical Status Classification, and the Charlson Comorbidity Index demonstrate only moderate predictive ability [1,2,5]. Criticisms of these tools include (1) their reliance on a limited subset of comorbidities or laboratory parameters, which may fail to capture real-world multimodal clinical information, and (2) dependence on International Classification of Diseases, 10th Revision coding, which may not capture the recent condition and severity of chronic disease in aging populations [4,6,7]. A broader range of current pre-operative data may therefore be necessary - similar to the integrative process used in routine clinical practice.

ChatGPT (Generative Pre-Training Transformer; OpenAI) is a widely used artificial intelligence (AI) tool capable of synthesizing complex information and responding to questions in real time [8,9]. The newest version, ChatGPT-5, demonstrates improved performance in combined image-text reasoning tasks in medicine, including radiology and oncology, and has outperformed GPT-4o in several clinical applications [10]. In addition, ChatGPT incorporates optical character recognition (OCR), enabling the extraction of linguistic information from medical images; however, its usefulness in medical data processing has not yet been fully evaluated [11].

If generative AI can appropriately integrate multiple pre-operative variables within a unified analytical framework, it may provide a new way to process complex clinical information in frail elderly patients. Therefore, this exploratory study aimed to (i) evaluate whether ChatGPT can provide clinically interpretable insights and (ii) assess the reproducibility of the generated outputs across two repeated evaluations. We further explored whether this framework yields clinically interpretable stratification patterns between survivors and non-survivors at 2 months postoperatively. This study does not address operative versus non-operative decision-making, because the dataset included only surgically treated patients.

## Materials and Methods

### Data source

Patients with pertrochanteric fractures treated between January 2018 and April 2024 at two Kobe University-affiliated hospitals were retrospectively reviewed. Exclusion criteria were: (1) Conservative (non-operative) treatment, including extremely high-risk patients deemed unsuitable for surgery, and (2) High-energy trauma or pathological fractures, to avoid confounding effects on mortality.

Ethics approval was obtained for the use of anonymized clinical data processed with ChatGPT, and the requirement for individual informed consent was waived by the institutional review board on June 17, 2025 (approval number: 2025-057). Study information was posted on the hospital website. All procedures followed the Japanese Ministry of Health, Labor and Welfare 2024 guidelines on AI-based use of clinical data.

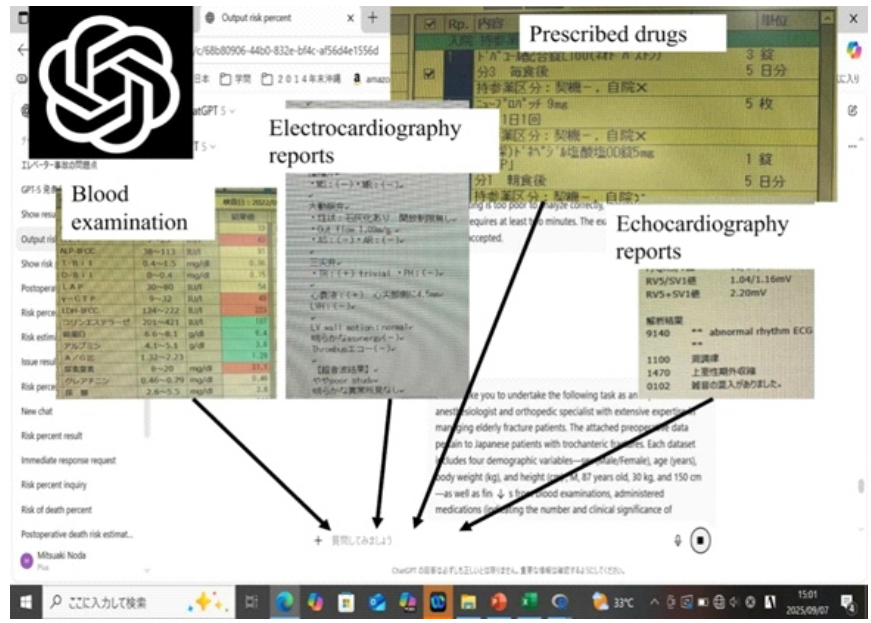


**Figure 1:** Schematic illustration of the ChatGPT analysis workflow. A predefined prompt was entered into ChatGPT together with manually entered demographic variables (sex, age, body weight, and height) and screenshot images of laboratory results, cardiac reports, and current medications.

### Pre-operative data collection

Routine pre-operative evaluation included

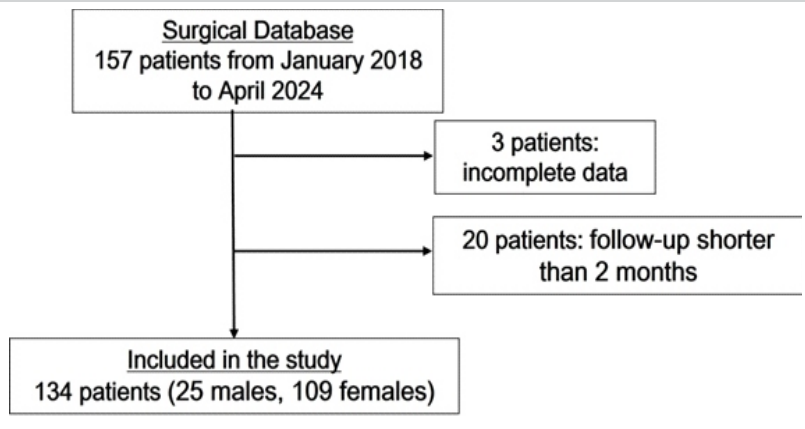


**Figure 2:** Flowchart of patient selection. Flowchart describing patient selection. Of 157 identified patients, 23 were excluded due to insufficient data (3 patients) or follow-up shorter than 2 months (20 patients). Therefore, 134 patients were included in the final analysis.

blood tests, an electrocardiogram (ECG), and an echocardiography. Laboratory data, cardiac reports, and medication lists were captured from the electronic medical record as de-identified screenshots. Demographic variables recorded were sex, age, body weight, and height (these items are not considered confidential under national AI guidelines).

### ChatGPT analysis

Data-sharing settings were configured to ensure uploaded content was not retained for third-party reuse. Image sets were attached to ChatGPT (GPT-4o and GPT-5). The identical prompts and data inputs were applied to each model twice, at least 1 week apart, with memory disabled to prevent previous outputs from influencing later sessions. To minimize variability, each analysis was performed in a new session with memory disabled (Fig. 1). Each uploaded dataset included four demographic variables – sex (Male/Female), age (years), body weight (kg), and height (cm) – as well as findings from blood examinations, reports of echocardiography and/or ECG, and administered medications (indicating the presence and clinical significance of comorbidities).

This study was not designed to validate predictive performance, but to examine the feasibility of multimodal data integration using a generative AI framework and the reproducibility of its outputs.

The standardized prompt instructed ChatGPT to integrate demographic and image-based clinical information and to return a single output value reflecting post-operative risk; the essential wording is summarized below:

"All patients underwent open reduction and internal fixation at a regional hospital with 100–250 beds in Japan.

Instructions:

1. Carefully estimate the 2-month post-operative risk of DEATH, exclusively-no explanations.

2. Incorporate the high-quality post-operative care available in Japan. A few surgically treated patients with similar clinical and demographic characteristics are expected to die within 2 months. Predictions must be evidence-based and realistically conservative.

3. Thoroughly evaluate the demographic data and all attached graphical images (labs, drugs, and results of ECG and/or echocardiography) by converting everything internally into a structured format (such as Excel-like tables). Then, accurate data extraction precedes clinical reasoning. Base your judgment on this structured interpretation. This process ensures accurate data extraction prior to clinical interpretation.

Ensure the evaluation is detailed, repeated, and reliable, you can use much time to complete.

All data are converted into a structured format before analysis.

Most patients underwent open reduction and internal fixation using either an intramedullary nail or a dynamic hip screw under the supervision of qualified anesthesiologists. Postoperatively, they were encouraged to mobilize early, generally with no restriction on weight bearing as part of the rehabilitation program. When patients had serious medical conditions, physicians from other departments provided additional support. Patients were classified as Alive or Dead

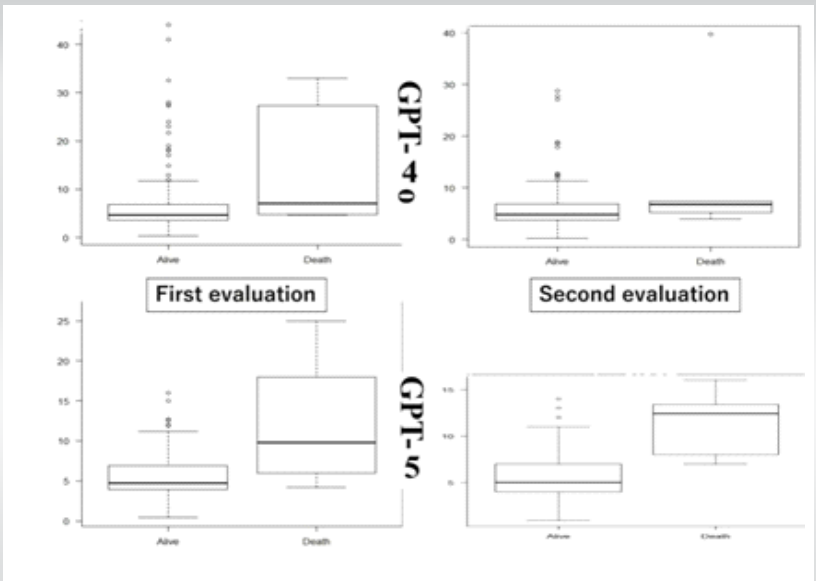


**Figure 3:** ChatGPT-generated mortality risk (%) values. Boxplots showing predicted mortality risk (%) generated by ChatGPT-4o and ChatGPT-5 in the Alive and Death groups for two separate evaluations (first and second). Box height represents the interquartile range (IQR), the central line indicates the median, and whiskers extend to 1.5 × IQR. Statistical significance was observed only in GPT-5 ($P < 0.05$).

| Table 1: Demographic characteristics of the alive and death groups at 2 months postoperatively | | | |
|---|---|---|---|
| | Alive ($n$=129) | Death ($n$=5) | $P$-value |
| Sex (M:F) | 1.031944444 | 02:03 | 0.233† |
| Age* | 87 (53–103) | 90 (81–97) | 0.438‡ |
| Body weight (kg)* | 44.7 (27–85) | 40.9 (33.7–50.3) | 0.331‡ |
| Height (cm)* | 151.2 (128–175) | 149.2 (143–161) | 0.484‡ |
| **This table summarizes patient characteristics (sex, age, body weight, and height) in the Alive and Death groups. *Mean and range are shown. †$P$-values for differences in sex were calculated using the Chi-square test. ‡$P$-values for differences in age, body weight, and height were calculated using Welch's $t$-test** | | | |

based on their survival status at 2 months after surgery.

### Statistical analysis

Analyses were conducted using EZR (R-based GUI). A P < 0.05 indicated statistical significance.

Group comparisons of sex, age, weight, and height were performed using Chi-square tests for sex and Welch's t-test or Mann–Whitney U-test as appropriate for continuous variables.

ChatGPT-derived output values (%) were also compared between groups using appropriate parametric or non-parametric testing. Given the exploratory nature of this study and the marked class imbalance, statistical significance in these comparisons was interpreted cautiously and was not used to infer predictive performance.

Agreement between repeated model outputs was assessed using Bland-Altman analysis [12].

## Results

### Demographic data

From the initial 157 patients, 23 were excluded due to incomplete pre-operative data (3 patients) or follow-up shorter than 2 months (20 patients). Thus, 134 patients were included in the study. The retrospective cohort consisted of 25 males and 109 females, with a mean age of 87 years (range, 53–103 years) (Fig. 2).

### Demographic comparison between alive group and death group

The fractures were classified into the alive group or the death group according to survival status within the 2-month post-operative period. The Alive group included 129 patients (23 males and 106 females; mean age 87 years, range 53–103 years), whereas the Death group comprised 5 patients (2 males and 3 females; mean age 90 years, range 81–97 years). There were no statistically significant differences between the two groups in gender, age, body weight, or height, although there was a markedly severe imbalance in sample size (129 vs. 5) (Table 1).

### Between-group separation of generated output values

ChatGPT-4o showed no significant difference in generated output risk between the two groups (P = 0.058 for the first evaluation and 0.184 for the second evaluation). In contrast, ChatGPT-5 demonstrated higher generated output values in the Death group compared with the Alive group: 4.7% vs. 9.8% in the first trial (P = 0.03), and 5.0% vs. 12.4% in the second trial (P = 0.002), respectively (Fig. 3).

### Agreement between repeated assessments in each GPT version

Bland-Altman analysis was performed to evaluate agreement between repeated mortality-risk predictions generated by ChatGPT-4o and ChatGPT-5. For ChatGPT-4o, the mean difference (bias) was 1.28% (95% confidence interval [CI], 0.09–2.47), indicating a small but statistically significant fixed bias. In contrast, ChatGPT-5 showed a mean difference of 0.18% (95% CI, −0.41–0.76), which was not statistically significant, suggesting the absence of systematic bias.

The limits of agreement (LoA) were markedly narrower for ChatGPT-5 (−5.53 to 5.88) than for ChatGPT-4o (−10.36 to 12.91), with corresponding LoA widths of approximately 11.4% and 23.3%, respectively (Table 2).

| Table 2: Bland-Altman analysis of agreement between repeated mortality risk predictions generated by ChatGPT-4o and ChatGPT-5 | | |
|---|---|---|
| Model | ChatGPT-4o | ChatGPT-5 |
| Mean difference (bias) | 1.28 (95% CI: 0.09–2.47) | 0.18 (95% CI: −0.41–0.76) |
| Limits of agreement | −10.36 to 12.91 | −5.53 to 5.88 |
| Width of limits of agreement (%) | 23.3 | 11.4 |
| **This table summarizes the mean bias and 95% limits of agreement for each model. CI: Confidence interval** | | |

## Discussion

Generative AI models, ChatGPT-4o and GPT-5, were used as multimodal analytical interfaces for integrating pre-operative clinical information in patients with pertrochanteric fractures treated at regional hospitals in Japan. Snapshot pre-operative information, including blood test results, cardiac findings, and prescribed medications, was entered into these models and was expected to be internally converted into structured data before analysis. The observed between-group separation, particularly with GPT-5, suggests that ChatGPT may integrate heterogeneous clinical information into outputs that reflect structured patterns consistent with clinical stratification, rather than serving as a direct predictive tool. This finding should not be interpreted as evidence of predictive accuracy. However, this behavior may resemble the integrative reasoning process used by clinicians when synthesizing multiple sources of clinical information in real-world decision-making.

### Technical advancement of GPT-5

There are several aspects in which GPT-5 appears superior to GPT-4, particularly regarding OCR capability and response consistency. With respect to OCR, a recent web-based evaluation reported acceptable GPT-5 performance across 10 document-understanding/OCR tasks, including retrieval of information from partially damaged receipts [13], although the model still struggled with tasks such as detecting small cracks on glass edges or accurately counting specified objects. In mammography analysis, GPT-5 consistently outperformed GPT-4o, achieving the highest scores among GPT variants in classifying density, distortion, masses, calcifications, and malignancy [14]. In the present study, random checks of OCR accuracy confirmed correct extraction of numerical and text-based medical data before analysis, particularly with GPT-5. Further validation using larger samples will be needed to confirm graphical-analysis advantages within our framework. Regarding consistency, GPT-5 also appeared to demonstrate greater reliability. Bland-Altman analysis in this study showed GPT-5 demonstrated a smaller mean bias and narrower LoA than GPT-4o. Taken together, GPT-5 appeared to produce more consistent and stable mortality-risk estimates within this dataset, despite the lack of statistical significance. Although generative AI systems are known to occasionally produce different responses to identical prompts on different occasions, this response variability inherent to generative AI systems may be gradually improving in GPT-5 [15]. Nevertheless, additional research is clearly required, given the still-limited number of published studies. These observations support the technical stability of GPT-5 within this framework, rather than confirming its clinical predictive superiority.

### Features of the analyzing model

The present study has several distinctive characteristics in its analytical framework. First, the model incorporated a wide range of routine pre-operative evaluations, including laboratory findings, cardiovascular assessments, and prescribed medications, reflecting real-world clinical information used in pre-operative decision-making. Previous studies have emphasized that renal function, diabetes, liver and respiratory disease, and prior myocardial infarction all influence early mortality risk after hip-fracture surgery, despite incomplete representation in traditional prognostic scoring systems [4,7,16]. The Charlson Comorbidity Index quantifies only a set number of comorbidities and does not reflect the severity of chronic conditions, which is instead indirectly captured through the prescribed medication profile used in our platform [6]. To reduce selection bias in risk scoring, the information base must be inclusive [7]. Second, this analytical framework allows expansion or modification of the dataset as needed [16]. Mortality risk varies according to institutional resources and geographic environment [2,7]. The prompt can therefore be adapted to incorporate region-specific or nation-specific clinical variables. However, this approach may not perform as effectively in developing countries, where clinical data are not consistently digitized or available online [17,18]. Third, compared with conventional AI systems, the present method is generally simple and efficient, with potentially a low workload for clinicians. This may help support timely surgery, which has been associated with improved outcomes [5]. Surgeons can complete the entire process – from data capture to AI-assisted estimation – within several minutes. Fourth, concerning cardiac data, we used interpreted reports prepared by automated systems or trained technicians, rather than raw cardiac images. Although ChatGPT may eventually be able to analyze medical images directly, written reports were considered a safer and more appropriate input format at this stage. Fifth, the attached data reflect the current comorbid health status, including changes in chronic disease burden in aging individuals. Because the number and severity of chronic conditions may change over time, particularly in elderly populations, real-time medical data are important for surgical risk estimation [6].

### Limitations of the current study system

This study has several limitations. First, the extreme imbalance in sample size between the two groups is critical. As a result, the marked class imbalance may have exaggerated the apparent between-group separation and limits the generalizability of the findings. These findings should be interpreted as hypothesis-generating rather than confirmatory, given the limited number

of death events. Second, because the present work focused exclusively on data obtained immediately after hospital admission, some clinically relevant factors were not included. Examples include dementia and pulmonary function testing, the latter often being unreliable when performed in bedridden patients or in those experiencing temporary cognitive disturbance shortly after hospitalization [7,16]. Third, the study could not compare outcomes directly with conventional mortality scoring systems in the same patients – such as the Nottingham Hip Fracture Score, ASA-PS, or the Charlson Comorbidity Index – due to insufficient background data [4,19]. For this reason, we intentionally avoided claiming that our method is superior to existing scoring systems. Fourth, the findings may not be fully applicable to all patients presenting to emergency departments, because extremely ill patients who were not selected for surgery were excluded from this database. Fifth, the prompt included contextual information such as the relatively low post-operative mortality rate in Japan. This instruction may have influenced the generated output values, acting as both a potential bias and a practical safeguard. Future work will remove this element to evaluate model behavior without this constraint.

### Future direction

We anticipate that ChatGPT will stimulate numerous innovative developments in orthopedics by enabling new approaches beyond conventional knowledge frameworks. Many surgeons recognize that AI capabilities are continually advancing. Hirosawa et al. [20] demonstrated how diverse medical images – including clinical photographs and diagnostic imaging – can be incorporated into AI systems. In line with this evolving approach, our study integrated multiple forms of pre-operative data. Future studies may expand further to include dynamic information, such as videos of gait or speech, because these factors have also been reported to influence medical risk and prognosis [5]. Validation using external datasets and resampling techniques will be essential.

## Conclusion

This study suggests that ChatGPT, particularly GPT-5, may enable exploratory separation between survivors and non-survivors after multimodal integration of pre-operative clinical information in pertrochanteric fracture patients. This should not be interpreted as proof of predictive validity. Rather, these findings indicate that a generative AI–assisted multimodal approach may provide clinically meaningful insights and support medical reasoning in orthopedic practice.

### Clinical Message

ChatGPT-assisted multimodal integration of routine pre-operative data may help synthesize complex clinical information in elderly hip fracture patients. This approach demonstrated reproducible separation between survivors and non-survivors, particularly with GPT-5. These findings should not be interpreted as predictive accuracy due to the exploratory design and marked class imbalance; rather, they reflect a structured form of clinical reasoning based on heterogeneous data integration. AI-assisted multimodal assessment may serve as a practical prototype to support complex decision-making in daily orthopedic practice.

**Declaration of patient consent:** The authors certify that they have obtained all appropriate patient consent forms. In the form, the patient has given the consent for his/ her images and other clinical information to be reported in the journal. The patient understands that his/ her names and initials will not be published and due efforts will be made to conceal their identity, but anonymity cannot be guaranteed.

**Conflict of interest:** Nil **Source of support:** None

## References

1. Lou'i Al-Husinat L, Azzam S, Sharie SA, Al Hseinat L, Araydah M, Al Modanat Z, et al. Impact of the American society of anesthesiologists (ASA) classification on hip fracture surgery outcomes: Insights from a retrospective analysis. BMC Anesthesiol 2024;24:271.

2. De Jong H, De Haan E, Van Rijckevorsel VA, Kuijper TM, Roukema GR. Multidimensional approach for predicting 30-day mortality in patients with a hip fracture: Development and external validation of the Rotterdam hip fracture mortality prediction-30 Days (RHMP-30). J Bone Joint Surg Am 2025;107:459-68.

3. Sofu H, Ucpunar H, Camurcu Y, Duman S, Konya MN, Gürsu S, et al. Predictive factors for early hospital readmission and 1-year mortality in elder patients following surgical treatment of a hip fracture. Ulus Travma Acil Cerrahi Derg 2017;23:245-50.

4. De Haan E, Roukema GR, Van Rijckevorsel VA, Kuijper TM, De Jong L, Dutch Hip Fracture Registry Collaboration. Risk factors for 30-days mortality after proximal femoral fracture surgery, a cohort study. Clin Interv Aging

2024;21:539-49.

5. Khan MA, Hossain FS, Ahmed I, Muthukumar N, Mohsen A. Predictors of early mortality after hip fracture surgery. Int Orthop 2013;37:119-24.

6. Ek S, Meyer AC, Hedstrom M, Modig K. Comorbidity and the association with 1-year mortality in hip fracture patients: Can the ASA score and the Charlson comorbidity index be used interchangeably? Aging Clin Exp Res 2022;34:129-36.

7. Sheikh HQ, Hossain FS, Aqil A, Akinbamijo B, Mushtaq V, Kapoor H. A comprehensive analysis of the causes and predictors of 30-day mortality following hip fracture surgery. Clin Orthop Surg 2017;9:10-8.

8. Alessi MR, Gomes HA, De Castro ML, Okamoto CT. Performance of ChatGPT in solving questions from the progress test (Brazilian national medical exam): A potential artificial intelligence tool in medical practice. Cureus 2024;16:e64924.

9. Noda M, Takahara S, Hayashi S, Inui A, Oe K, Matsushita T. Evaluating ChatGPT's performance in classifying pertrochanteric fractures based on arbeitsgemeinschaft fur osteosynthesefragen/orthopedic trauma association (AO/OTA) standards. Cureus 2025;17:e78068.

10. Benchmarking GPT-5 for Zero-Shot Multimodal Medical Reasoning in Radiology and Radiation Oncology. Available from: https://openai.com/ja-JP/index/introducing-gpt-5 [Last accessed on 2025 Dec 25].

11. Posner KM, Bakus C, Basralian G, Chester G, Zeiman M, O'Malley GR, et al. Evaluating ChatGPT's capabilities on orthopedic training examinations: An analysis of new image processing features. Cureus 2024;16:e55945.

12. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.

13. GPT-5 for Vision: Results from 80+ Real-World Tests; 2025. Available from: https://blog.roboflow.com/gpt-5-vision-multimodal-evaluation [Last accessed on 2025 Dec 25].

14. GPT-5 Demonstrates Mammography VQA Performance on BI-RADS Assessment and Malignancy Classification; 2025. Available from: https://quantumzeitgeist.com/gpt-5-demonstrates-mammography-vqa-performance-on-bi-rads-assessment-and-malignancy-classification [Last accessed on 2025 Dec 25].

15. Zhou Y, Moon C, Szatkowski J, Moore D, Stevens J. Evaluating ChatGPT responses in the context of a 53-year-old male with a femoral neck fracture: A qualitative analysis. Eur J Orthop Surg Traumatol 2024;34:927-55.

16. Gundel O, Thygesen LC, Gogenur I, Ekeloef S. Postoperative mortality after a hip fracture over a 15-year period in Denmark: A national register study. Acta Orthop 2020;91:58-62.

17. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health 2023;2:e0000198.

18. Gao CA, Howard FM, Markov NS, Dyer EC, Ramesh S, Luo Y, et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. NPJ Digit Med 2023;6:75.

19. Hasan O, Barkat R, Rabbani A, Rabbani U, Mahmood F, Noordin S. Charlson comorbidity index predicts postoperative complications in surgically treated hip fracture patients in a tertiary care hospital: Retrospective cohort of 1045 patients. Int J Surg 2020;82:116-20.

20. Hirosawa T, Harada Y, Tokumasu K, Ito T, Suzuki T, Shimizu T. Evaluating ChatGPT-4's diagnostic accuracy: Impact of visual data integration. JMIR Med Inform 2024;12:e55627.

**Conflict of Interest:** Nil
**Source of Support:** Nil

**Consent:** The authors confirm that informed consent was obtained from the patient for publication of this article

**How to Cite this Article**

Noda M, Takahara S, Hayashi S, Inui A, Oe K, Matsushita T. Clinical Implications of ChatGPT-Assisted Multimodal Pre-operative Assessment in Elderly Pertrochanteric Fracture Patients: An Exploratory Study. Journal of Orthopaedic Case Reports 2026 June;16(06):220-226.