Available online at www.medicinescience.org

ORIGINAL ARTICLE

Medicine Science International Medical Journal

Medicine Science 2026;15(3):1171-6

How do artificial intelligence chatbots address patient concerns about neuraxial anesthesia?

Fatih Olus, Huseyin Babun

Akdeniz University, Faculty of Dentistry, Department of Oral and Maxillofacial Surgery, Antalya, Türkiye

Received 05 February 2026; Accepted 05 April 2026
Available online 13 August 2026 with doi: 10.5455/medscience.2026.02.033

Abstract

Artificial intelligence (AI) chatbots are increasingly used by patients seeking medical information before surgery; however, the clinical appropriateness and educational value of chatbot-generated responses regarding neuraxial anesthesia remain uncertain. This study evaluated how three widely accessible AI chatbots—ChatGPT, Gemini, and Copilot—address common patient questions about spinal and epidural anesthesia. Thirty frequently asked questions were developed and grouped into four domains: general information and procedure, safety and risks, intraoperative experience, and postoperative recovery. Each question was submitted identically to the freely accessible, publicly available versions of the three chatbots. Ten board-certified anesthesiologists, blinded to the chatbot source, independently rated each response using a 5-point Likert scale (1 = very inappropriate, 5 = very appropriate). In total, 900 ratings (30 questions × 3 chatbots × 10 evaluators) were analyzed. ChatGPT achieved the highest overall mean score (4.66 ± 0.49), followed by Gemini (4.41 ± 0.49) and Copilot (3.32 ± 0.61). Between-platform differences were statistically significant in a linear mixed-effects model accounting for clustering by rater and question (Wald $\chi^2(2) = 1173.44$, $p < 0.001$). Model-based pairwise comparisons demonstrated that ChatGPT scored higher than Gemini and Copilot, and Gemini scored higher than Copilot (all $p < 0.001$). Across all four domains, ChatGPT and Gemini consistently outperformed Copilot; ChatGPT ranked highest in three domains, while Gemini showed slightly better performance for intraoperative experience–related questions. In this expert-rating study conducted at a single time point, ChatGPT-generated responses were rated as more clinically appropriate than those produced by Gemini and Copilot. These findings suggest that AI chatbots may support preoperative patient education, but their outputs require cautious interpretation, ongoing expert oversight, and periodic re-evaluation as platforms evolve.

Keywords: Artificial intelligence, neuraxial anesthesia, spinal anesthesia, epidural anesthesia, patient education, preoperative care

Introduction

Neuraxial anesthesia, including spinal and epidural techniques, is widely used for lower abdominal, pelvic, obstetric, and lower-extremity surgeries due to its efficacy, rapid onset, and favorable safety profile [1-3]. Spinal anesthesia produces a dense and predictable block through intrathecal local anesthetic injection, whereas epidural anesthesia involves catheter placement in the epidural space, allowing titratable dosing and extended analgesia when needed [2,3]. Although both modalities act within the neuroaxis, they differ significantly in pharmacologic characteristics, duration, and clinical indications.

Despite their extensive use, patients often possess limited understanding of neuraxial anesthesia, leading to misconceptions and heightened anxiety during the preoperative period. Several large cohort studies have shown that severe complications such as permanent neurologic injury or paralysis are extremely rare, with estimated incidences ranging from 1:10,000 to 1:50,000 depending on patient population and comorbidities [4-6]. Transient events such as hypotension, post–dural puncture headache, urinary retention, and temporary paresthesias remain the most commonly encountered side effects and are usually self-limited [7,8]. Nevertheless, patients routinely overestimate these risks, particularly in relation to spinal cord injury and paralysis.

CITATION
Olus F, Babun H. How do artificial intelligence chatbots address patient concerns about neuraxial anesthesia?. Med Science. 2026;15(3):1171-6.

Corresponding Author: Huseyin Babun, Akdeniz University, Faculty of Dentistry, Department of Oral and Maxillofacial Surgery, Antalya, Türkiye
Email: huseyinbabun@gmail.com

DOI: 10.5455/medscience.2026.02.033

International studies consistently show that patient knowledge regarding anesthesia—particularly neuraxial techniques—is often limited**.** Many patients are unsure about the differences between spinal and epidural anesthesia and remain unaware of the actual risks, benefits, and expected intraoperative experience [9,10]. Systematic reviews further indicate that misunderstandings about the role of the anesthesiologist, needle placement, intraoperative awareness, and loss of control contribute significantly to preoperative anxiety [11,12]. Cross-national survey data reveal that misconceptions about paralysis and neurologic injury are among the most common false beliefs, even in countries with well-established anesthesia services [13].

Preoperative anxiety is strongly influenced by these informational gaps. Randomized trials have demonstrated that providing structured, high-quality educational materials significantly reduces perioperative anxiety and improves patient satisfaction, particularly for individuals undergoing regional anesthesia [14]. In the absence of reliable educational resources, many patients turn to online information sources, where accuracy and consistency vary widely. Misinformation can exacerbate existing fears, leading to reluctance or refusal of neuraxial anesthesia even when clinically appropriate. Patients increasingly turn to artificial intelligence (AI) chatbots as part of this online search for medical guidance, particularly when timely access to anesthesiologists or structured educational materials is limited.

The rapid emergence of AI chatbots—such as ChatGPT, Gemini, and Copilot—has transformed how patients seek medical information. These large language model (LLM)-based tools offer immediate, conversational answers and are increasingly used as alternative sources of health information. However, existing evidence suggests that while AI chatbots can produce helpful general explanations, their medical accuracy, completeness, and reliability remain inconsistent [15]. This variability may be particularly concerning in specialized fields such as neuraxial anesthesia, where misunderstandings can amplify patient anxiety.

Given the widespread use of AI chatbots and the critical importance of accurate preoperative counseling, it is essential to examine how effectively these platforms address common patient concerns about spinal and epidural anesthesia. Therefore, the aim of this study was to evaluate the accuracy, completeness, and educational value of responses generated by three freely accessible, publicly available versions of AI chatbots—ChatGPT, Gemini, and Copilot—when answering frequently asked patient questions about spinal and epidural anesthesia.

Material and Methods

Study Design and Setting

This cross-sectional, evaluator-blinded study was conducted to assess the accuracy and appropriateness of responses generated by three freely accessible, publicly available versions of AI chatbots regarding spinal and epidural anesthesia. The study was performed between March and August 2025. This study was designed and reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cross-sectional studies [16].

Development of the Question Set

A total of 30 frequently asked questions (FAQs) related to spinal and epidural anesthesia were developed by two board-certified anesthesiologists with clinical experience in neuraxial anesthesia. The questions reflected common concerns expressed by patients during routine preoperative consultations.

Item generation was informed by institutional preoperative information materials, commonly addressed topics in anesthesia society patient information resources, and recurring themes encountered during routine clinical practice related to patient concerns and misconceptions. Although no formal psychometric validation was performed, the questions underwent iterative expert review to ensure content clarity, clinical relevance, and coverage of key counseling elements, including indications, expected benefits, alternative anesthesia options, common transient side effects, rare but serious complications, contraindications, intraoperative experience, and postoperative recovery expectations. For structured evaluation, the questions were categorized into four domains established prior to data collection:

1. General Information and Procedure,
2. Safety and Risks,
3. Intraoperative Experience,
4. Postoperative Recovery.

Together, these four domains were designed to reflect the core informational dimensions typically addressed during preoperative counseling for neuraxial anesthesia. The “General Information and Procedure” domain focused on fundamental explanations of technique and indications. The “Safety and Risks” domain addressed common concerns regarding complications and contraindications. The “Intraoperative Experience” domain explored patient expectations during surgery under neuraxial anesthesia, while the “Postoperative Recovery” domain examined anticipated recovery patterns and transient postoperative effects. This structured framework allowed for a comprehensive and clinically meaningful evaluation of chatbot-generated responses.

To enhance representativeness, the question set was intentionally designed to mirror common themes encountered during routine preoperative consultations. The included items reflected frequently asked patient concerns regarding procedural understanding, complication risk, intraoperative awareness, and postoperative recovery. Care was taken to balance straightforward explanatory questions with more anxiety-driven or risk-focused inquiries, thereby capturing varying levels of informational complexity. Although no formal validation process was undertaken, this structured and clinically grounded approach aimed to ensure that the questions were broadly representative of typical patient concerns in neuraxial anesthesia practice.

DOI: 10.5455/medscience.2026.02.033

AI Chatbot Platforms and Response Generation

Each question was entered separately and identically into three widely used AI chatbots—ChatGPT (OpenAI, GPT-4), Gemini (Google), and Copilot (Microsoft). All platforms were accessed in May 2025 using their freely accessible, publicly available versions to reflect information available to the public. All chatbot interactions were conducted via the web-based public interfaces of each platform.

The exact patient-facing question text was used for all queries, without any additional prompting constraints such as role specification, formatting instructions, reading-level adjustments, requests for conciseness, or bullet-point formatting. The full list of the 30 questions and their predefined domain assignments are provided in Supplementary Appendix A.

To avoid contextual bias, each query was submitted in a new chat session, with chat history, personalization, memory functions, and any "helpful mode" features disabled or not utilized where applicable. Web-browsing and citation-generation features were not enabled during response generation.

Each question was submitted once per platform, reflecting a real-world scenario in which patients typically rely on a single chatbot-generated response rather than repeated queries.

Blinding Procedure

To minimize evaluator bias, all chatbot-generated responses were de-identified. Platform names, stylistic identifiers, and metadata were removed. Each response was assigned a randomized alphanumeric code. For every question, the three chatbot-generated responses were randomly ordered using a computer-generated sequence. Each evaluator received a unique coded response booklet without any indication of chatbot identity. Evaluators were instructed not to discuss the rating process with one another. After completion of all blinded ratings, chatbot identities were unblinded for reporting purposes, and platform-specific example responses are presented in Supplementary Appendix B.

Evaluation by Anesthesiologists

A panel of 10 board-certified anesthesiologists, each with a minimum of five years of clinical experience, independently rated all chatbot-generated answers. Evaluators were blinded to the source of each response. Each answer was assessed using a 5-point Likert scale evaluating overall appropriateness:

1 = Very inappropriate

2 = Inappropriate

3 = Neutral / partially appropriate

4 = Appropriate

5 = Very appropriate

Evaluators were instructed to base their overall appropriateness ratings on their expert clinical judgment, considering factors such as medical accuracy, completeness of information, alignment with current safety standards and guidelines, clarity of explanation, and the quality of risk communication. Although a single global score was assigned for each response, raters were encouraged to integrate these dimensions holistically when determining their rating.

A total of 900 evaluations were generated (30 questions × 3 chatbots × 10 raters).

Statistical Analysis

Data analysis was performed using IBM SPSS Statistics (version 25, IBM Corp., Armonk, NY, USA) and Python. Continuous variables were presented as mean ± standard deviation (SD). Because ratings were repeated within raters and clustered within questions, between-platform differences in expert-rated clinical appropriateness were analyzed using a linear mixed-effects model with chatbot as a fixed effect and random intercepts for rater and question. Pairwise between-platform contrasts were performed with multiplicity control. A two-sided p value < 0.05 was considered statistically significant. This approach was selected to appropriately account for the clustered and repeated-measures structure of the data.

Ethical Considerations

This study did not involve human participants, patient data, or identifiable personal information. All chatbot outputs were generated from publicly accessible AI platforms using predefined hypothetical patient questions. No interaction with real patients occurred, and no clinical data was collected. Therefore, ethical committee approval and informed consent were not required in accordance with institutional and national research regulations.

Results

Each individual expert rating constituted the unit of analysis. A total of 900 individual evaluations were collected (30 questions × 3 chatbots × 10 raters). Descriptive statistics for each platform are shown in Table 1.

Table 1. Descriptive statistics of expert-rated clinical appropriateness scores by platform

Platform	Mean ± SD	Min	Max
ChatGPT	4.66 ± 0.49	3	5
Gemini	4.41 ± 0.49	4	5
Copilot	3.32 ± 0.61	1	5

Min and Max values represent the lowest and highest individual appropriateness scores assigned by expert raters across all evaluated chatbot responses

Overall Appropriateness Scores

Each chatbot contributed 300 expert ratings (30 questions × 10 raters). Among the evaluated platforms, ChatGPT achieved the highest overall rating (mean 4.66 ± 0.49), followed by Gemini (mean 4.41 ± 0.49). Copilot demonstrated substantially lower performance (mean 3.32 ± 0.61). These descriptive findings are summarized in Table 1.

DOI: 10.5455/medscience.2026.02.033

Between-Platform Statistical Comparison

Between-platform differences in expert-rated clinical appropriateness were assessed using a linear mixed-effects model accounting for clustering within raters and questions. The overall effect of chatbot was statistically significant (Wald χ^2 (2) = 1173.44, $p < 0.001$). Model estimated mean appropriateness scores were highest for ChatGPT (4.66 ± 0.49), followed by Gemini (4.41 ± 0.49) and Copilot (3.32 ± 0.61). Model-based pairwise comparisons indicated that ChatGPT scored significantly higher than Gemini ($\Delta = 0.247$, $p < 0.001$) and Copilot ($\Delta = 1.343$, $p < 0.001$), and Gemini scored significantly higher than Copilot ($\Delta = 1.097$, $p < 0.001$).

Domain-Based Comparisons

When questions were grouped into four predefined domains—General Information, Safety and Risks, Intraoperative Experience, and Postoperative Recovery—ChatGPT and Gemini consistently demonstrated high scores across all categories, whereas Copilot yielded lower ratings in every domain. This descriptive pattern is illustrated in Figure 1, which compares domain-level mean scores across the three platforms. ChatGPT performed best in General Information and Safety, while Gemini showed slightly higher performance in the Intraoperative domain. Copilot exhibited its lowest scores in Safety and Risks and Intraoperative Experience.

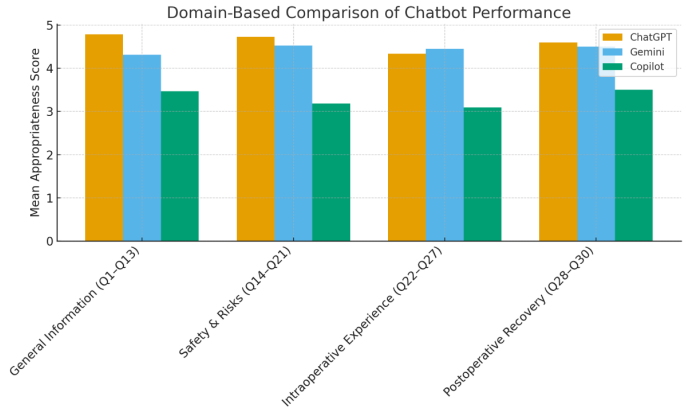

Figure 1. Comparison of mean appropriateness scores across four question domains (General Information, Safety and Risks, Intraoperative Experience, Postoperative Recovery) for ChatGPT, Gemini, and Copilot. Error bars represent standard deviations

Per-Question Performance Trends

Figure 2 presents a question-by-question comparison of mean scores for all 30 items. ChatGPT maintained stable, high-level performance across questions with minimal fluctuation. Gemini also demonstrated a stable pattern with intermediate scores. Copilot showed considerably greater variability, including several questions with mean scores below 3.0, particularly in questions involving intraoperative sensations and risk communication. These per-item trends highlight systematic performance differences between platforms and reinforce the results of the domain-based analysis.

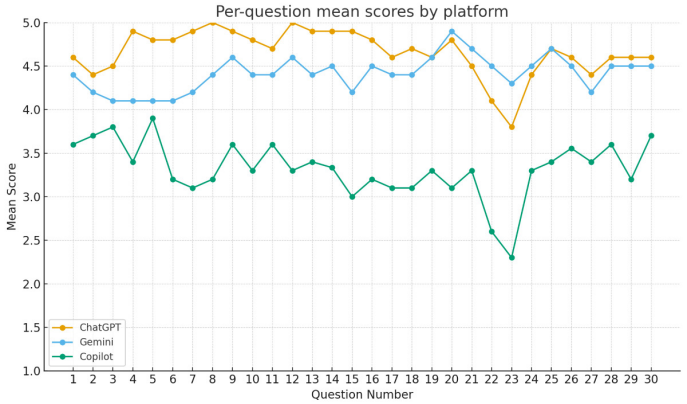

Figure 2. Question-by-question comparison of mean appropriateness scores for ChatGPT, Gemini, and Copilot across all 30 patient questions about spinal and epidural anesthesia

Discussion

This study compared the expert-rated clinical appropriateness of responses generated by three publicly accessible AI chatbots—ChatGPT, Gemini, and Copilot—regarding frequently asked questions about spinal and epidural anesthesia. A significant performance gradient was observed: ChatGPT consistently outperformed both Gemini and Copilot, demonstrating higher clinical appropriateness, completeness, and clinical relevance across nearly all question domains.

Comparison With Existing Anesthesia Literature

Although spinal and epidural anesthesia techniques are generally safe and associated with very low rates of severe neurological complications [4-6], patients often hold exaggerated fears about paralysis, permanent nerve damage, or intraoperative pain [7,12,13]. Misconceptions related to neuraxial anesthesia can contribute to heightened preoperative anxiety, which is already prevalent among surgical patients [10]. Accurate patient education is therefore essential.

Our findings support earlier evidence showing that newer LLMs—particularly GPT-4—produce more clinically aligned content than prior versions. Choi et al. reported that ChatGPT-4 generated higher-quality anesthesia explanations than ChatGPT-3.5 [17]. Similarly, Jin et al. demonstrated that GPT-4 can produce structured, accurate educational content for anesthesiology residents, although hallucinations and reference errors may still occur [18]. These observations align with our results, where ChatGPT showed the most complete and clinically reasonable responses.

Gemini demonstrated intermediate performance. While generally accurate, its answers were occasionally less detailed in risk-related questions, consistent with prior findings that some models provide reduced depth in safety-critical topics [17]. Interestingly, Gemini performed slightly better than ChatGPT in questions related to intraoperative sensory experience, possibly reflecting model differences in descriptive language generation.

DOI: 10.5455/medscience.2026.02.033

Copilot showed the lowest performance and greatest variability. Several responses lacked essential risk information or provided oversimplified explanations. Similar limitations have been reported in other domains, where models produced incomplete, outdated, or unverifiable anesthesia content [17]. This reinforces the need for cautious clinical oversight when chatbots are used in patient-facing contexts.

Clinical Implications and Patient Communication

High-quality information can meaningfully reduce perioperative anxiety and improve patient satisfaction [10,11,14]. ChatGPT and Gemini—especially ChatGPT—appear capable of providing reliable, easy-to-understand explanations that may complement clinician counseling. Their high performance in domains such as general information and postoperative expectations is consistent with what patients commonly seek before anesthesia [7,10-12].

However, Copilot's performance highlights how inconsistent or incomplete AI explanations may reinforce misconceptions rather than correct them. Given that patient misunderstanding of anesthesia is common [12,13], reliance on lower-performing models without verification could be problematic.

From a practical standpoint, AI chatbots could be incorporated as adjunct educational tools rather than stand-alone counseling resources. For example, hospitals may recommend specific AI platforms to provide preliminary, standardized information before the preoperative anesthesia visit, allowing patients to arrive with more structured questions. Chatbots may also be used to reinforce explanations after consultation, particularly for frequently asked questions about risks, sensory experience, and recovery expectations. However, their use should remain supervised, with clinicians clearly informing patients that chatbot-generated information does not replace individualized medical assessment. Institutions considering integration should establish guidance regarding appropriate platforms, version monitoring, and periodic content auditing to ensure ongoing clinical accuracy.

Alignment With Broader AI Chatbot Research

Our findings are in line with broader evaluations of medical chatbots showing that LLMs have meaningful educational potential but remain imperfect. Prior studies have noted issues such as factual drift, inconsistent depth, and occasional citation errors in ChatGPT-generated medical content [17,18]. Model-to-model variability is also well documented, with differences in architecture and training data affecting accuracy and risk communication [17,18]. Even high-performing systems may produce confident yet incorrect statements, as demonstrated by Goodman et al. [15]. These patterns parallel our results, reinforcing that although chatbots can supplement patient education, clinical oversight remains essential.

Strengths and Limitations

This study has several notable strengths, including the use of freely accessible, publicly available chatbot versions that reflect real-world patient access, as well as a clinically relevant question set designed around common patient concerns. The blinded evaluation by ten anesthesiologists further strengthened the reliability of the ratings by ensuring expert judgment and minimizing potential bias. A mixed-effects modeling approach was used to account for clustering of ratings within raters and questions.

However, certain limitations should also be acknowledged. Furthermore, although the question set was developed by experienced anesthesiologists and designed to reflect common patient concerns, it did not undergo formal psychometric validation. No structured pilot testing, construct validation, or reliability analysis of the questionnaire was performed. Therefore, the findings should be interpreted within the context of this methodological limitation.

The analysis was restricted to English-language responses, and chatbot performance may vary across other languages or cultural contexts. AI models are continually updated, and comparative performance may evolve over time as underlying architectures and safety alignment mechanisms are modified. Therefore, the present findings reflect model behavior at a specific evaluation point and should not be interpreted as fixed performance hierarchies. Additionally, LLM outputs are inherently stochastic, and repeated prompting under identical conditions may yield variable responses. In the present study, each question was submitted once per platform to reflect a typical real-world patient interaction. However, this approach does not capture potential within-model variability across repeated queries. Future studies could employ repeated sampling designs or temperature-controlled prompting to quantify response stability and explore how variability may influence clinical appropriateness ratings. Such analyses would provide a more nuanced understanding of model consistency over time and across iterations.

Future Research

Future research should expand this work by evaluating chatbot performance across multiple languages, exploring their integration as structured adjuncts within preoperative education pathways, and conducting longitudinal assessments to understand how model updates affect accuracy over time. Additional efforts should focus on developing anesthesia-specific fine-tuned language models capable of delivering more precise and context-aware information. Finally, studies examining patient perceptions, comprehension, and anxiety levels after receiving chatbot-generated explanations would provide valuable insights into the real-world impact of these tools on preoperative experience.

The ethical and legal implications of integrating AI chatbots into patient education must also be carefully considered.

DOI: 10.5455/medscience.2026.02.033

Chatbots generate probabilistic responses that may vary over time and may not always align with current clinical guidelines. Unlike licensed healthcare professionals, AI systems do not assume legal responsibility for the information they provide. Therefore, reliance on chatbot-generated medical advice without professional verification may raise concerns regarding misinformation, liability, and patient safety. Transparency about the limitations of AI tools, clear communication that they do not replace individualized medical consultation, and institutional oversight mechanisms are essential to ensure ethically responsible implementation in clinical settings.

Conclusion

In this expert-rating study conducted at a single time point, ChatGPT-generated responses were rated as more clinically appropriate than those produced by Gemini and Copilot for the evaluated patient questions on spinal and epidural anesthesia. Given the dynamic nature of AI model development, these comparative findings represent platform performance at the time of evaluation and may evolve with future model updates. These findings align with growing evidence that advanced LLMs can support—but not replace—clinical communication. Responsible implementation will require continued validation, transparency, and expert oversight.

Ethical Approval

This study did not involve human subjects or patient data and therefore did not require approval from an ethics committee. All anesthesiologists who participated as expert evaluators provided voluntary consent prior to data collection.

Patient Consent

Informed consent was not required for this study, as it did not include patients, case reports, or any identifiable human data.

Conflict of Interest

The authors confirm the absence of any conflicts associated with this work.

Funding

The authors declare that they have received no financial support for the study.

Acknowledgments

The authors sincerely thank the anesthesiologists who participated as expert raters for their valuable insights and contributions.

References

1. Brown DL. Spinal, epidural, and caudal anesthesia. In: Miller RD, ed. Miller's Anesthesia. 7th ed. Churchill Livingstone Elsevier, Philadelphia, 2010;1611-38.
2. Barash PG, Cullen BF, Stoelting RK, et al. Clinical Anesthesia. 7th ed. Lippincott Williams & Wilkins, Philadelphia, 2013.
3. Neal JM, Barrington MJ, Brull R, et al. The second ASRA practice advisory on neurologic complications associated with regional anesthesia and pain medicine: executive summary 2015. Reg Anesth Pain Med. 2015;40:401-30.
4. Moen V, Dahlgren N, Irestedt L. Severe neurological complications after central neuraxial blockades in Sweden 1990-1999. Anesthesiology. 2004;101:950-9.
5. Auroy Y, Narchi P, Messiah A, et al. Serious complications related to regional anesthesia: results of a prospective survey in France. Anesthesiology. 1997;87:479-86.
6. Auroy Y, Benhamou D, Bargues L, et al. Major complications of regional anesthesia in France: the SOS Regional Anesthesia Hotline Service. Anesthesiology. 2002;97:1274-80.
7. Turnbull DK, Shepherd DB. Post-dural puncture headache: pathogenesis, prevention and treatment. Br J Anaesth. 2003;91:718-29.
8. Pozza DH, Tavares I, Cruz CD, Fonseca S. Spinal cord injury and complications related to neuraxial anaesthesia procedures: a systematic review. Int J Mol Sci. 2023;24:4665.
9. Nitzani D, Nicholls J, Maslowski K, et al. Patient perception of consent processes for epidural analgesia in induction of labour: a qualitative study. Anaesthesia. 2025;80:1199-206.
10. Bello CM, Eisler P, Heidegger T. Perioperative anxiety: current status and future perspectives. J Clin Med. 2025;14:1422.
11. Lee A, Chui PT, Gin T. Educating patients about anesthesia: a systematic review of randomized controlled trials of media-based interventions. Anesth Analg. 2003;96:1424-31.
12. Leite F, da Silva LM, Biancolin SE, et al. Patient perceptions about anesthesia and anesthesiologists before and after surgical procedures. Sao Paulo Med J. 2011;129:224-9.
13. Mavridou P, Dimitriou V, Manataki A, et al. Patient's anxiety and fear of anesthesia: effect of gender, age, education, and previous experience of anesthesia. A survey of 400 patients. J Anesth. 2013;27:104-8.
14. Jlala H, French J, Foxall G, et al. Effect of preoperative multimedia information on perioperative anxiety in patients undergoing procedures under regional anaesthesia. Br J Anaesth. 2010;104:369-74.
15. Goodman RS, Patrinely JR, Stone CA Jr, et al. Accuracy and reliability of chatbot responses to physician questions. JAMA Netw Open. 2023;6:e2336483.
16. Cuschieri S. The STROBE guidelines. Saudi J Anaesth. 2019;13:S31-4.
17. Choi J, Oh AR, Park J, et al. Evaluation of the quality and quantity of artificial intelligence-generated responses about anesthesia and surgery: using ChatGPT 3.5 and 4.0. Front Med (Lausanne). 2024;11:1400153.
18. Jin Z, Abola R, Bargnes V 3rd, et al. The utility of generative artificial intelligence chatbot (ChatGPT) in generating teaching and learning material for anesthesiology residents. Front Artif Intell. 2025;8:1582096.