



Original Article

Corresponding Author

Samuel K. Cho

<https://orcid.org/0000-0001-7511-2486>

Department of Orthopedic Surgery, Icahn School of Medicine at Mount Sinai, 787 11th Avenue, 7th Floor, New York, NY 10019, USA
Email: samuel.cho@mountsinai.org

Received: December 10, 2023

Revised: January 29, 2024

Accepted: February 17, 2024

See the commentary on “Performance of a Large Language Model in the Generation of Clinical Guidelines for Antibiotic Prophylaxis in Spine Surgery” via <https://doi.org/10.14245/ns.2448236.118>.



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2024 by the Korean Spinal Neurosurgery Society

INTRODUCTION

Chat generative pre-trained transformer (ChatGPT) is a 175 billion-parameter large language model (LLM) developed by OpenAI and trained on a large, nondomain specific corpus of textual data from the Internet.¹ Since ChatGPT's public release in November 2022, the model has gained widespread attention for its “human-like” responses to textual prompts, including its apparent ability to show deductive reasoning skills, coherence,

Performance of a Large Language Model in the Generation of Clinical Guidelines for Antibiotic Prophylaxis in Spine Surgery

Bashar Zaidat, Nancy Shrestha, Ashley M. Rosenberg, Wasil Ahmed, Rami Rajjoub, Timothy Hoang, Mateo Restrepo Mejia, Akiro H. Duey, Justin E. Tang, Jun S. Kim, Samuel K. Cho

Department of Orthopedic Surgery, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Objective: Large language models, such as chat generative pre-trained transformer (ChatGPT), have great potential for streamlining medical processes and assisting physicians in clinical decision-making. This study aimed to assess the potential of ChatGPT's 2 models (GPT-3.5 and GPT-4.0) to support clinical decision-making by comparing its responses for antibiotic prophylaxis in spine surgery to accepted clinical guidelines.

Methods: ChatGPT models were prompted with questions from the North American Spine Society (NASS) Evidence-based Clinical Guidelines for Multidisciplinary Spine Care for Antibiotic Prophylaxis in Spine Surgery (2013). Its responses were then compared and assessed for accuracy.

Results: Of the 16 NASS guideline questions concerning antibiotic prophylaxis, 10 responses (62.5%) were accurate in ChatGPT's GPT-3.5 model and 13 (81%) were accurate in GPT-4.0. Twenty-five percent of GPT-3.5 answers were deemed as overly confident while 62.5% of GPT-4.0 answers directly used the NASS guideline as evidence for its response.

Conclusion: ChatGPT demonstrated an impressive ability to accurately answer clinical questions. GPT-3.5 model's performance was limited by its tendency to give overly confident responses and its inability to identify the most significant elements in its responses. GPT-4.0 model's responses had higher accuracy and cited the NASS guideline as direct evidence many times. While GPT-4.0 is still far from perfect, it has shown an exceptional ability to extract the most relevant research available compared to GPT-3.5. Thus, while ChatGPT has shown far-reaching potential, scrutiny should still be exercised regarding its clinical use at this time.

Keywords: Artificial intelligence, Antibiotic prophylaxis, Orthopedic surgery

and continuity in its generated responses.

Artificial intelligence (AI) models have long been used in the medical field to assist in imaging analysis and provide clinical decision-making support. Recently, natural language processing (NLP) models have been applied to many medical tasks, including billing code generation,^{2,3} prediction of patient outcomes and mortality,⁴ and prediction of emergency hospital admissions.⁵ NLP chatbots further have the potential to be applied in patient-facing primary care settings.⁶ Researchers have begun investi-

gating novel applications of ChatGPT for various medical tasks, such as the generation of accurate and clinically relevant medical information and guidance. Because the model is capable of passing the United States Medical Licensing Exam (USMLE),^{7,8} some researchers have suggested that ChatGPT may assist in medical education and clinical decision-making.

To assess the potential of ChatGPT as a clinical decision-making support tool, it is imperative to assess not only whether ChatGPT can give accurate recommendations, but also if its reasoning aligns with established clinical guidelines that follow evidence-based practices. Generating clinical guidelines is a labor-intensive process that requires continuous updates; therefore, applying ChatGPT reliably in this area could prove beneficial to medical practitioners if it can accurately synthesize patient scenarios and provide relevant recommendations. One study found that ChatGPT could appropriately understand hypothetical patient infection scenarios and recognized complex management considerations, however, it failed to distinguish clinically important factors without prior prompting and sometimes failed to give appropriate management plans.⁹

ChatGPT currently has 2 models available to the public: GPT-3.5 and GPT-4.0. GPT-4.0 is a more advanced system with a monthly subscription of \$20, purported to be safer and more accurate than GPT-3.5, the free version, according to ChatGPT's developer OpenAI.¹ In this study, we assess ChatGPT's 2 models in their ability to generate accurate clinical guidelines on antibiotic prophylaxis in spine surgery using 16 questions from the North American Spine Society (NASS) Evidence-Based Clinical Guidelines for Multidisciplinary Spine Care.

MATERIALS AND METHODS

The NASS's Evidence-Based Clinical Guidelines for Multidisciplinary Spine Care: *Antibiotic Prophylaxis in Spine Surgery*,¹⁰ revised in 2013, was used as the clinical standard. Two ChatGPT models, GPT-3.5 and GPT-4.0, were asked independently to answer each of the 16 clinical questions addressed in this guideline—an example is shown in Fig. 1. A new chat was created for each question to avoid biases from prior prompts. If the question did not include a reference to spine surgery, it was modified to include these words. Otherwise, the prompts were presented to the models exactly as they appeared in the NASS guidelines. Each response was recorded verbatim, summarized, and compared to the NASS guideline recommendations for accuracy. Acknowledging that GPT has a tendency to provide long-winded and sometimes repetitive information in its responses,

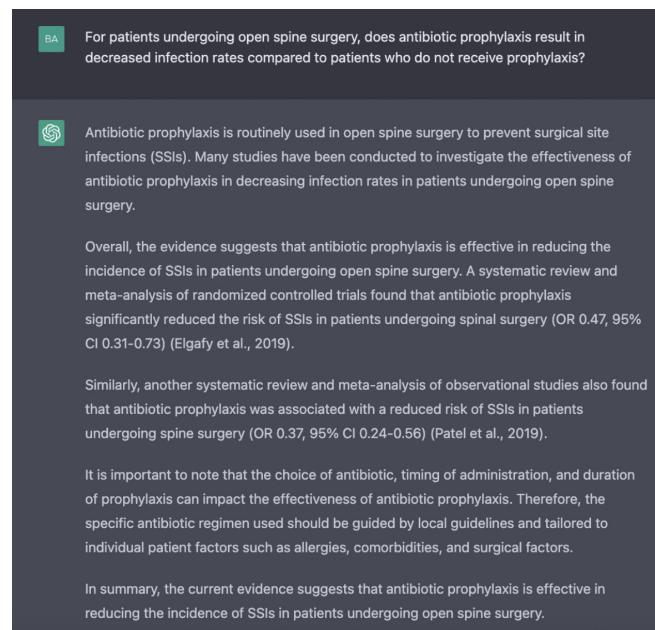


Fig. 1. Example response of ChatGPT's prompt and response in the GPT-3.5 model. ChatGPT, chat generative pre-trained transformer.

the responses were summarized to the key recommendations, which were compared to the NASS recommendations to determine accuracy. In cases where NASS guidelines gave no recommendation due to insufficient evidence, we judged the GPT response correct if it was also able to conclude that no specific recommendation could be provided. Additionally, to account for potential outdated answers in the NASS guidelines, we searched the literature for studies published in the last ten years to discern if the NASS guidelines were still supported by recent research and determine if ChatGPT responses were unaligned with the NASS guidelines but aligned with more recent research.

An example of this process is shown in Fig. 2. This response was graded as accurate. ChatGPT produced the correct response that prophylactic antibiotics reduce surgical site infection (SSI) incidence and stated that studies have shown no difference in single-dose versus multiple-dose prophylaxis regimens in spine surgeries. This corresponds well with the NASS guideline recommendation that preoperative antibiotic prophylaxis has shown decreased infection rates.

Responses for both the GPT-3.5 and GPT-4.0 were evaluated for accuracy compared to the NASS guidelines, and the performance of the 2 models were compared. To further assess the advancement from GPT-3.5 to 4.0, we additionally evaluated 2 aspects of GPT-4.0's advancements that have been highlighted by the OpenAI team: a decrease in overconfidence, which was

Table 1. ChatGPT's performance (GPT-3.5 & GPT-4.0) compared to the North American Spine Society (NASS) clinical guidelines for clinical questions relating to antibiotic prophylaxis in spine surgery, broken down by question category

Antibiotic prophylaxis clinical question category	GPT-3.5			GPT-4.0		
	Accurate	Inaccurate	Overconfident	Accurate	Inaccurate	Cited NASS
All questions (n = 16)	10 (62.5)	6 (37.5)	4 (25)	13 (81)	3 (19)	10 (62.5)
Efficacy (n = 4)	3 (75)	1 (25)	1 (25)	4 (100)	0 (0)	3 (75)
Protocol (n = 4)	1 (25)	3 (75)	3 (75)	3 (75)	1 (25)	3 (75)
Redosing (n = 1)	1 (100)	0 (0)	1 (100)	1 (100)	0 (0)	1 (100)
Discontinuation(n = 1)	0 (0)	1 (100)	-	0 (0)	1 (100)	-
Wound drains (n = 1)	1 (100)	0 (0)	-	1 (100)	0 (0)	1 (100)
Body habitus (n = 1)	1 (100)	0 (0)	-	1 (100)	0 (0)	1 (100)
Comorbidities (n = 2)	2 (100)	0 (0)	-	1 (50)	1 (50)	-
Complications (n = 2)	1 (50)	1 (50)	-	2 (100)	0 (0)	1 (50)

Values are presented as number (%).

ChatGPT, chat generative pre-trained transformer.

defined as when the model would support statements that had conflicting data in the literature and guidelines without referencing the ambiguity in the literature, and a more advanced capability to provide real citations. Thus, we assessed GPT-3.5's tendency for overconfidence in its responses as well as GPT-4.0's ability to reference the NASS guidelines.

RESULTS

1. Question Content

Sixteen questions in the NASS guidelines addressed the use of antibiotic prophylaxis in open spine surgery. Four of these regarded the efficacy of antibiotic prophylaxis in preventing infections in open spine surgery; 4 others are concerned with protocols for selection, intraoperative, and postoperative application of antibiotic prophylaxis; 1 is on redosing; 2 are on discontinuation; 3 discuss the effects of comorbidities and body habitus; and 2 are on complications arising from antibiotic prophylaxis. Six of the total questions posed binary responses, while the 10 remaining contained more complex responses.

2. Overall Performance

Comparing ChatGPT's GPT-3.5 model with the NASS clinical guideline recommendations, 10 responses (62.5%) were accurate and 6 (37.5%) were inaccurate. The model adequately answered all questions related to antibiotic redosing, wound drains, body habitus and comorbidities. It was inaccurate for 1 of 4 questions on antibiotic efficacy, 3 of 4 questions on antibiotic protocol, 1 of 1 question on antibiotic discontinuation, and 1 of 2 questions on antibiotic complications. Compared to the

NASS clinical guideline recommendations, ChatGPT's GPT-4.0 model correctly answered 13 responses (81%) and inaccurately answered 3 (19%). The more advanced model answered all questions related to antibiotic efficacy, redosing, wound drains, body habitus and complications. It inaccurately answered 1 of 4 questions on antibiotic protocol, 1 of 1 question on antibiotic discontinuation, and 1 of 2 questions on antibiotic comorbidities (Tables 1, 2; Fig. 3).

DISCUSSION

1. Efficacy

Three of 4 efficacy questions were binary questions concerned the efficacy of antibiotics in preventing SSIs in general open spine surgery, spine surgery with implants, and spine surgery without implants. Both ChatGPT models adequately answered that antibiotic prophylaxis decreases the rate of SSIs in all 3 cases, with GPT-4.0 even citing the NASS guidelines as evidence for its responses. Studies have shown a positive effect of antibiotic prophylaxis on reducing SSIs in general spine surgery as well as instrumented and non-instrumented spine surgery¹¹⁻¹³ which aligns with NASS guidelines.

The last question concerned the expected rate of SSIs when antibiotic prophylaxis is applied. The NASS guidelines report an infection rate of 0.7%–10% while a recent meta-analysis reported an infection rate of 3.1% in all spine surgeries with prophylaxis.¹⁴ GPT-3.5 failed to report a prevalence of SSIs, instead falsely citing that the incidence of SSIs is lowered by 60%–80% with antibiotic prophylaxis use. Additionally, the guidelines report that patients with diabetes face an increased risk of SSIs,

Table 2. A comparison of the clinical questions and North American Spine Society (NASS) guidelines for antibiotics prophylaxis in spine surgery with corresponding ChatGPT model recommendations (GPT-3.5 & GPT-4.0)

Category	NASS question	NASS guidelines	GPT-3.5 recommendation	GPT-4.0 recommendation
Efficacy	For patients undergoing open spine surgery, does antibiotic prophylaxis result in decreased infection rates compared to patients who do not receive prophylaxis?	Preoperative prophylactic antibiotics are suggested to decrease infection rates in patients undergoing spine surgery. Grade of recommendation: B For a typical, uncomplicated lumbar laminotomy and discectomy, a single preoperative dose of antibiotics is suggested to decrease the risk of infection and/or discitis. Grade of recommendation: B	The use of prophylactic antibiotics for patients undergoing open spine surgery has been shown to significantly reduce the incidence of surgical site infections (SSIs) compared to patients who do not receive prophylaxis, and it is considered standard practice.	In summary, antibiotic prophylaxis is beneficial in reducing infection rates in patients undergoing open spine surgery.
Efficacy	For patients undergoing open spine surgery <i>without</i> spinal implants, does antibiotic prophylaxis result in decreased infection rates compared to patients who do not receive prophylaxis?	Prophylactic antibiotics are suggested to decrease the rate of spinal infections following uninstrumented lumbar spinal surgery. Grade of recommendation: B	Evidence suggests that antibiotic prophylaxis can reduce the risk of SSIs in patients undergoing open spine surgery without spinal implants. However, the use of antibiotics should be balanced against the risk of developing antibiotic resistance and other adverse effects associated with their use.	In summary, prophylactic antibiotics are indeed beneficial in reducing the rate of spinal infections after surgeries that do not involve the use of instruments in the lumbar region. These findings underscore the importance of antibiotic prophylaxis in reducing infection risks in open spine surgeries without implants, although more research is needed to determine the most effective agents, doses, and administration routes.
Efficacy	For patients undergoing open spine surgery <i>with</i> spinal implants, does antibiotic prophylaxis result in decreased infection rates as compared to patients who do not receive prophylaxis?	Prophylactic antibiotics may be considered to decrease the rate of infections following instrumented spine fusion. Grade of recommendation: C	Studies suggest that antibiotic prophylaxis can reduce the incidence of SSIs in patients undergoing open spine surgery with spinal implants. The specific antibiotic regimen and duration of prophylaxis may vary based on several factors.	In summary, antibiotic prophylaxis can reduce infection rates in spine surgery, but the effectiveness can vary based on the type of surgery, the patient's condition, and the duration of the antibiotic administration. It is important to consider the individual circumstances of each patient when determining the best prophylactic strategy.
Efficacy	What rate of SSIs can be expected with the use of antibiotic prophylaxis, considering both patients with and patients without medical comorbidities?	Consensus statement: Despite appropriate prophylaxis, the rate of SSIs in spine surgery is 0.7%–10%. The expected rate for patients without comorbidities ranges from 0.7%–4.3% and for patients with comorbidities ranges from 2.0%–10%. Current best practice with antibiotic protocols has failed to eliminate (reach an infection rate of 0.0%) SSIs. Despite appropriate prophylaxis, diabetes carries an increased infection rate compared with non-diabetic patients. Level of evidence: III There is insufficient evidence to make a statement regarding the impact of obesity on the rate of surgical site infection in prophylaxed patients. Level of evidence: I (Insufficient)	The rate of SSIs can vary depending on various factors such as the type of surgery, patient characteristics, and the quality of the surgical facility. Patients with comorbidities including diabetes, obesity, and immune deficiencies may be at a higher risk of developing SSIs. However, the use of antibiotic prophylaxis can significantly reduce the risk of SSIs by as much as 60%–80%.	The rate of SSIs with the use of antibiotic prophylaxis varies depending on several factors, including wound class, the presence of comorbid conditions, the duration of the surgical procedure, the timing of antibiotic administration, and whether prophylaxis is used at all. In a study conducted at Desse Referral Hospital, an overall SSI rate of 23.4% was observed among patients who underwent major surgery. It was noted that 48.6% of these cases involved inappropriate timing of prophylactic antibiotic administration, which is a significant factor associated with the risk of SSIs. For a more comprehensive understanding and additional data on this subject, further research from multiple sources would be needed to compare results across different patient populations and healthcare settings.

(Continued)

Table 2. A comparison of the clinical questions and North American Spine Society (NASS) guidelines for antibiotics prophylaxis in spine surgery with corresponding ChatGPT model recommendations (GPT-3.5 & GPT-4.0) (Continued)

Category	NASS question	NASS guidelines	GPT-3.5 recommendation	GPT-4.0 recommendation
Protocol	For patients receiving antibiotic prophylaxis prior to open spine surgery, what are the recommended drugs, their dosages, administration routes and timing resulting in decreased postoperative infection rates?	<p>Preoperative antibiotic prophylaxis is suggested to decrease infection rates in patients undergoing spine surgery. In typical, uncomplicated spinal procedures, the superiority of one agent, dose or route of administration over any other has not been clearly demonstrated. When determining the appropriate drug choice, the patient's risk factors, allergies, length and complexity of the procedure and issues of antibiotic resistance should be considered. Grade of recommendation: B</p> <p>In typical, uncomplicated spinal procedures, a single dose of preoperative prophylactic antibiotics with intraoperative redosing as needed is suggested. Grade of recommendation: B</p> <p>Consensus statement: In patients with comorbidities or for those undergoing complicated spine surgery, alternative prophylactic regimens including redosing, gram-negative coverage or the addition of intrawound application of vancomycin or gentamicin, are suggested to decrease the incidence of SSI's when compared to standard prophylaxis regimens.</p> <p>Consensus statement: Comorbidities and risk factors reviewed in the literature include obesity, diabetes, neurologic deficits, incontinence, preoperative serum glucose level of > 125 mg/dL or a postoperative serum glucose level of > 200 mg/dL, trauma and prolonged multilevel instrumented surgery. Olsen et al. (2003)¹⁴ and Olsen et al. (2008)¹⁵ are provided as support studies below to further define the risk factors associated with surgical site infection in spine surgery patients.</p>	<p>The choice of antibiotic prophylaxis for open spine surgery should be based on the specific needs of the patient and the potential risks of infection. However, here are some general recommendations for antibiotic prophylaxis prior to open spine surgery, but note that patient factors such as allergies, comorbidities, and previous infections should be considered when selecting appropriate antibiotic prophylaxis:</p> <p>Cefazolin: Cefazolin is the most commonly used antibiotic for prophylaxis in spine surgery. The recommended dosage is 1–2 grams administered intravenously (IV) within 60 minutes prior to surgery.</p> <p>Vancomycin: Vancomycin is an alternative to cefazolin in patients with a beta-lactam allergy or in areas with a high prevalence of methicillin-resistant <i>Staphylococcus aureus</i> (MRSA). The recommended dosage is 15–20 mg/kg administered IV within 60 minutes prior to surgery.</p> <p>Clindamycin: Clindamycin is another alternative to cefazolin in patients with a beta-lactam allergy. The recommended dosage is 600–900 mg administered IV within 60 minutes prior to surgery.</p> <p>The timing of administration should be within 60 minutes prior to surgery to ensure adequate tissue levels at the time of incision. In addition, the duration of prophylaxis should not exceed 24 hours.</p>	<p>The guidelines from the NASS for antibiotic prophylaxis in spine surgery suggest the following:</p> <p>Preoperative Prophylaxis: A single preoperative dose of antibiotics is suggested to decrease infection rates in patients undergoing spine surgery, including both uninstrumented and instrumented spinal surgeries. The choice of antibiotic should consider the patient's risk factors, allergies, the length and complexity of the surgery, and antibiotic resistance concerns.</p> <p>Drug choice and administration: No specific agent, dose, or route of administration has been shown to be superior to others in uncomplicated cases. Broad-spectrum antibiotics may be considered for patients with risk factors for polymicrobial infection.</p> <p>Redosing: Intraoperative redosing within 3 to 4 hours may be considered to maintain therapeutic antibiotic levels throughout the procedure.</p> <p>Postoperative discontinuation: For simple cases, a single preoperative dose with intraoperative redosing as needed is suggested. Prolonged regimens may be considered for complex cases or patients with comorbidities such as diabetes, obesity, or neuromuscular disease.</p> <p>It's important for healthcare providers to tailor antibiotic prophylaxis to the individual patient and the specifics of the surgical procedure.</p>

(Continued)

Table 2. A comparison of the clinical questions and North American Spine Society (NASS) guidelines for antibiotics prophylaxis in spine surgery with corresponding ChatGPT model recommendations (GPT-3.5 & GPT-4.0) (Continued)

Category	NASS question	NASS guidelines	GPT-3.5 recommendation	GPT-4.0 recommendation
Protocol	For patients receiving antibiotic prophylaxis prior to open spine surgery without spinal implants, what are the recommended drugs, their dosages, administration routes and timing resulting in decreased postoperative infections rates?	Preoperative antibiotic prophylaxis is suggested to decrease infection rates in patients undergoing spine surgery without spinal implants. In these typical, uncomplicated spinal procedures, the superiority of one agent, dose or route of administration over any other has not been clearly demonstrated. When determining the appropriate drug choice, the patient's risk factors, allergies, length and complexity of the procedure and issues of antibiotic resistance should be considered. Grade of recommendation: B In typical, uncomplicated open spine surgery without spinal implants, a single dose of preoperative prophylactic antibiotics with intraoperative redosing as needed is suggested. Grade of recommendation: B	The choice of antibiotic prophylaxis for open spine surgery without spinal implants should be based on the most common pathogens in the surgical site and the patient's individual risk factors. The following are the general recommendations: Cefazolin is the recommended first-line antibiotic prophylaxis. It should be administered intravenously at a dose of 1–2 grams 30 minutes to 1 hour before surgery. The infusion should be completed within 1 hour. Vancomycin should be used if the patient has a known allergy to cephalosporins. Its dosage is 15 mg/kg, 1 hour before surgery. For patients with beta-lactam allergy, clindamycin or aztreonam can be used as an alternative. Clindamycin should be administered intravenously at a dose of 600–900 mg and aztreonam at a dose of 1–2 grams 30 minutes to 1 hour before surgery.	For patients undergoing open spine surgery without spinal implants, preoperative antibiotic prophylaxis is recommended to decrease infection rates. The guidelines suggest that a single dose of preoperative prophylactic antibiotics, with intraoperative redosing as needed, is adequate for typical uncomplicated cases. Intraoperative redosing should be considered within 3 to 4 hours to maintain therapeutic levels throughout the procedure. The specific agent, dose, and route of administration have not shown any clear superiority over others. Therefore, it is suggested that the choice of antibiotic should be tailored to the patient's risk factors, allergies, length and complexity of the procedure, and concerns about antibiotic resistance.
Protocol	For patients receiving antibiotic prophylaxis prior to open spine surgery with spinal implants, what are the recommended drugs, their dosages, administration routes and timing resulting in decreased postoperative infections rates?	Preoperative antibiotic prophylaxis is suggested to decrease infection rates in patients undergoing spine surgery with spinal implants. In these complex spinal procedures, the superiority of one agent, dose or route of administration over any other has not been clearly demonstrated. When determining the appropriate drug choice, the patient's risk factors, allergies, length and complexity of the procedure and issues of antibiotic resistance should be considered. Grade of recommendation: B Preoperative antibiotic prophylaxis is recommended to decrease infection rates in patients undergoing spine surgery with spinal implants. The North American Spine Society suggests: For simple, uncomplicated spine surgery without instrumentation or comorbidities: a single preoperative dose of the antibiotic of choice with intraoperative redosing as needed. For instrumented spine surgery, prolonged procedures, or patients with comorbidities such as diabetes, neuromuscular disease, spinal cord injury, or general spine trauma: a single preoperative dose of the antibiotic of choice is suggested with consideration for additional gram-negative coverage and/or the application of intrawound vancomycin or gentamicin. However, there is no clear superiority of one agent, dose, or route of administration over another, and the specific antibiotic choice should consider patient risk factors, allergies, procedure length and complexity, and antibiotic resistance issues.	The choice of antibiotics, dosages, administration routes, and timing for antibiotic prophylaxis prior to open spine surgery with spinal implants can vary based on factors like the patient's medical history, the type of surgical procedure, and the surgeon's preference. General guidelines are: Cefazolin (1–2 g IV) is a first-line antibiotic and should be administered within 60 minutes prior to incision. Vancomycin (15–20 mg/kg IV) can be added for patients with a history of MRSA colonization, recent MRSA infection, or beta-lactam allergy. It should be administered within 2 hours prior to incision.	Preoperative antibiotic prophylaxis is recommended to decrease infection rates in patients undergoing spine surgery with spinal implants. The North American Spine Society suggests: For simple, uncomplicated spine surgery without instrumentation or comorbidities: a single preoperative dose of the antibiotic of choice with intraoperative redosing as needed. For instrumented spine surgery, prolonged procedures, or patients with comorbidities such as diabetes, neuromuscular disease, spinal cord injury, or general spine trauma: a single preoperative dose of the antibiotic of choice is suggested with consideration for additional gram-negative coverage and/or the application of intrawound vancomycin or gentamicin. However, there is no clear superiority of one agent, dose, or route of administration over another, and the specific antibiotic choice should consider patient risk factors, allergies, procedure length and complexity, and antibiotic resistance issues.

(Continued)

Table 2. A comparison of the clinical questions and North American Spine Society (NASS) guidelines for antibiotics prophylaxis in spine surgery with corresponding ChatGPT model recommendations (GPT-3.5 & GPT-4.0) (Continued)

Category	NASS question	NASS guidelines	GPT-3.5 recommendation	GPT-4.0 recommendation
Protocol	What is a reasonable algorithmic approach for antibiotic selection for a given patient?	<p>Consensus statement: Simple uncomplicated spine surgery (without instrumentation or comorbidities) => one single preoperative dose of antibiotic of choice with intraoperative redosing as needed.</p> <p>Consensus statement: Instrumented spine surgery, prolonged procedures, comorbidities (e.g., diabetes, neuromuscular disease, cord injury or general spine trauma) => one single preoperative dose of antibiotic of choice+consideration of additional gram-negative coverage and/or the application of intrawound vancomycin or gentamicin.</p>	<p>Antibiotic selection for a patient who just underwent spine surgery would depend on several factors. A general guide would be to (1) determine the patient's risk factors for SSI; (2) determine the type of surgical procedure, which will dictate the duration of antibiotic prophylaxis; (3) choose an appropriate antibiotic regimen based on the most common organisms that cause SSIs in spinal surgeries, which are usually Gram-positive cocci; (4) consider additional risk factors, which may require a longer duration of prophylactic antibiotics; and (5) monitor for adverse effects and adjust therapy accordingly.</p>	<p>For a patient who has undergone spine surgery, the approach to antibiotic selection would be more specific due to the surgical context and the potential for different infecting organisms. Here's an algorithmic approach tailored for such a scenario: (1) Determine the patient's risk of infection post-surgery, (2) Administer prophylactic antibiotics prior to incision, (3) Note any unusual intraoperative findings that may necessitate a deviation from standard prophylaxis, (4) Continue prophylactic antibiotics for a short duration postoperatively, (5) Vigilantly monitor for signs of SSI during the postoperative period, (6) If infection is suspected, obtain cultures from the wound or any drainage, (7) If an infection is suspected, start empiric broad-spectrum antibiotics that cover skin flora and other common organisms associated with SSIs, (8) Once culture and sensitivity results are available, tailor the antibiotic therapy to the specific organisms isolated and their sensitivities, (9) Consider the ability of the antibiotic to penetrate bone and the surgical site, (10) Take into account patient-specific factors such as renal and hepatic function, (11) For confirmed SSIs, the duration of therapy can be quite prolonged, (12) Follow antibiotic stewardship principles to prevent resistance, (13) Consider involving an infectious disease specialist, (14) Regularly reassess the patient's response to the antibiotic therapy, and (15) Plan for long-term follow-up.</p> <p>It's important to note that these steps are generalized and may need to be adapted based on the patient's individual circumstances and institutional protocols. Additionally, consultation with an infectious disease specialist is often beneficial in complex cases.</p>

(Continued)

Table 2. A comparison of the clinical questions and North American Spine Society (NASS) guidelines for antibiotics prophylaxis in spine surgery with corresponding ChatGPT model recommendations (GPT-3.5 & GPT-4.0) (Continued)

Category	NASS question	NASS guidelines	GPT-3.5 recommendation	GPT-4.0 recommendation
Redosing	For patients receiving antibiotic prophylaxis prior to open spine surgery, what are the intraoperative redosing recommendations for the recommended drugs (including dosages and time of administration) resulting in decreased postoperative infection rates?	Consensus statement: Intraoperative redosing within 3-4 hours may be considered to maintain therapeutic antibiotic levels throughout the procedure. The superiority of one drug has not been demonstrated in the literature. When determining the appropriate drug choice, the patient's risk factors, allergies, length and complexity of the procedure and issues of antibiotic resistance should be considered.	Redosing should occur at regular intervals during the surgical procedure to ensure that adequate concentrations of the antibiotic are present in the body too. The recommended redosing guidelines for the most common antibiotics are: Cefazolin: 1-2 grams IV at the time of induction of anesthesia. For surgeries lasting longer than 4 hours, redose with 1 gram IV every 3-4 hours. Vancomycin: 15-20 mg/kg IV at the time of induction of anesthesia. For surgeries lasting longer than 4 hours, redose with the same dose every 6 hours.	For patients undergoing open spine surgery, intraoperative redosing of antibiotics is recommended to maintain therapeutic levels throughout the procedure. Redosing within 3 to 4 hours may be considered for this purpose. However, there is no consensus on the superiority of one drug over another. The choice of the antibiotic should take into account the patient's risk factors, allergies, the length and complexity of the surgery, and antibiotic resistance issues. Furthermore, for patients with comorbidities or those undergoing complicated spine surgery, alternative prophylactic regimens are suggested to decrease the incidence of SSIs. However, there is insufficient evidence to recommend specific alternative regimens, though promising alternatives include redosing, adding gram-negative coverage, and the intrawound application of vancomycin or gentamicin.
Discontinuation	For patients receiving antibiotic prophylaxis prior to open spine surgery, does discontinuation of prophylaxis at 24 hours result in decreased or increased postoperative infection rates as compared to longer periods of administration?	For typical, uncomplicated cases, a single dose of preoperative prophylactic antibiotics with intraoperative redosing as needed is suggested to decrease the risk of surgical site infection. Grade of recommendation: B Prolonged postoperative regimens may be considered in complex situations (i.e., trauma, cord injury, neuromuscular disease, diabetes or other comorbidities). Comorbidities and complex situations reviewed in the literature include obesity, diabetes, neurologic deficits, incontinence, preoperative serum glucose level of >125 mg/dL or a postoperative serum glucose level of >200 mg/dL, trauma, prolonged multilevel instrumented surgery and other comorbidities. Grade of recommendation: C	The available evidence suggests that discontinuing prophylaxis at 24 hours does not result in increased postoperative infection rates compared to longer durations of administration. However, the optimal duration of prophylaxis may vary depending on the specific patient population, surgical technique, and other factors.	The study I found reports that discontinuation of antibiotic prophylaxis at 24 hours postoperatively does not result in increased postoperative infection rates compared to longer periods of administration in open spine surgery. A propensity-matched analysis of 4,454 patients, with 2,672 receiving 24 hours of postoperative antibiotics, showed no significant difference in infection rates compared to patients who did not receive postoperative antibiotics. The rates were 1.8% for patients who received antibiotics and 1.5% for those who did not. Additionally, prolonged antibiotic use was not associated with an increased risk of <i>Clostridium difficile</i> infection or infections with multidrug-resistant organisms. For a more comprehensive understanding, further research from multiple sources would be ideal to provide a thorough consensus on the topic.

(Continued)

Table 2. A comparison of the clinical questions and North American Spine Society (NASS) guidelines for antibiotics prophylaxis in spine surgery with corresponding ChatGPT model recommendations (GPT-3.5 & GPT-4.0) (Continued)

Category	NASS question	NASS guidelines	GPT-3.5 recommendation	GPT-4.0 recommendation
Wound drains	For patients receiving antibiotic prophylaxis prior to open spine surgery and who receive placement of wound drains at wound closure, does discontinuation of prophylaxis at 24 hours result in decreased or increased postoperative infection rates as compared to discontinuation of antibiotics at time of drain removal?	There is insufficient evidence to make a recommendation for or against the early discontinuation of antibiotic prophylaxis in patients with wound drains. Grade of recommendation: I (insufficient evidence) The use of drains is not recommended as a means to reduce infection rates following single level surgical procedures. Grade of recommendation: I (insufficient evidence)	The optimal duration of antibiotic prophylaxis for open spine surgery and the use of wound drains is still a matter of debate among medical professionals. However, some studies have suggested that there may be no significant difference in postoperative infection rates between patients who discontinue prophylaxis at 24 hours versus those who discontinue antibiotics at the time of drain removal.	The research regarding the discontinuation of antibiotic prophylaxis in open spine surgery for patients with wound drains varies. One propensity-matched analysis study found no significant difference in the rate of SSIs between patients who received 24 hours of postoperative antibiotics and those who did not. This study included a large sample size and controlled for confounding factors, suggesting that extended antibiotic use beyond 24 hours postoperatively does not decrease the risk of SSIs. The NASS Evidence-Based Clinical Guideline on Antibiotic Prophylaxis in Spine Surgery provides recommendations based on the highest quality clinical literature available as of December 2006. The recommendations in this guideline, although not specifying the exact duration, aim to assist in delivering optimum treatment to prevent SSIs. The guideline itself does not detail the outcomes for discontinuing antibiotics at 24 hours versus at the time of drain removal, but it emphasizes the importance of evidence-based protocols. In summary, the available evidence suggests that there is no significant benefit in continuing antibiotic prophylaxis beyond 24 hours postoperatively, even when wound drains are in place. It is important to note that guidelines and recommendations can evolve, and practices may vary based on individual patient circumstances, the type of surgery performed, and institutional protocols.

(Continued)

Table 2. A comparison of the clinical questions and North American Spine Society (NASS) guidelines for antibiotics prophylaxis in spine surgery with corresponding ChatGPT model recommendations (GPT-3.5 & GPT-4.0) (Continued)

Category	NASS question	NASS guidelines	GPT-3.5 recommendation	GPT-4.0 recommendation
Body habitus	For patients receiving antibiotic prophylaxis prior to spine surgery, should the recommended protocol differ based upon body habitus (e.g., body mass index)?	Obese patients are at higher risk for postoperative infection, when given a standardized dose of antibiotic prophylaxis. In spite of this conclusion, there is insufficient evidence to make a recommendation for or against recommending a different protocol for patients based upon body habitus. Grade of recommendation: I (insufficient evidence)	There is currently limited evidence to suggest that the recommended antibiotic prophylaxis protocol should differ based on a patient's body habitus, such as body mass index (BMI), prior to spine surgery. While some studies have suggested that higher doses of antibiotics may be necessary for patients with obesity or higher BMI due to altered pharmacokinetics, there is not enough evidence to support a different prophylaxis protocol based solely on BMI.	The protocol for antibiotic prophylaxis in spine surgery relative to a patient's body habitus, specifically BMI, does merit consideration given that bodyweight and BMI have been correlated with higher rates of SSI in various surgical procedures. The NASS Evidence-Based Clinical Guideline on Antibiotic Prophylaxis in Spine Surgery provides a comprehensive overview, it does not explicitly mention adjustments for body habitus or BMI. A publication from SICOT-1 indicates that while obese patients are at a higher risk for postoperative infection when given a standardized dose of antibiotic prophylaxis, there is insufficient evidence to make a recommendation for or against recommending a different protocol for patients based upon body habitus. This conclusion suggests that while the issue is recognized, the evidence at the time of that publication was not strong enough to inform a change in protocol. In summary, while there is recognition that obese patients may have a higher risk for infection with standardized dosing, the guidelines are not universally adjusted for BMI due to varying levels of evidence. Nonetheless, some studies support the idea of adjusting doses based on bodyweight to reduce the risk of SSIs.
Comorbidities	For patients receiving antibiotic prophylaxis prior to open spine surgery, do comorbidities (other than obesity) such as diabetes, smoking, nutritional depletion, immunodeficiencies and concurrent use of antithrombotic therapies alter the recommendations for antibiotic prophylaxis?	Consensus statement: In patients with comorbidities or for those undergoing complicated spine surgery, alternative prophylactic regimens are suggested to decrease the incidence of SSIs when compared to standard prophylaxis regimens. There is insufficient evidence to make a recommendation for or against the specific alternative regimens that are efficacious. However, promising alternative regimens that have been studied include redosing, gram-negative coverage and the addition of intrawound application of vancomycin or gentamicin. Grade of Recommendation: I (insufficient evidence)	Comorbidities such as diabetes, smoking, nutritional depletion, immunodeficiencies, and concurrent use of antithrombotic therapies can alter the recommendations for antibiotic prophylaxis in patients receiving prophylaxis prior to open spine surgery. In cases of patients with these comorbidities, it may be necessary to extend the duration of antibiotic prophylaxis, delay the initiation of prophylaxis, or adjust the dose of the prophylaxis.	The NASS Evidence-Based Clinical Guideline for Antibiotic Prophylaxis in Spine Surgery, while comprehensive in addressing efficacy and appropriateness of antibiotic prophylaxis protocols, repeat dosing, discontinuation, wound drains, and special considerations including comorbidities, does not appear to provide specific differentiated recommendations for patients with comorbidities such as diabetes, smoking, nutritional depletion, immunodeficiencies, and concurrent use of antithrombotic therapies directly in the summary accessible through PubMed.

(Continued)

Table 2. A comparison of the clinical questions and North American Spine Society (NASS) guidelines for antibiotics prophylaxis in spine surgery with corresponding ChatGPT model recommendations (GPT-3.5 & GPT-4.0) (Continued)

Category	NASS question	NASS guidelines	GPT-3.5 recommendation	GPT-4.0 recommendation
Comorbidities	For patients with a history of MRSA infection, does prophylaxis with vancomycin reduce infections with MRSA compared to other antimicrobial agents?	There is insufficient evidence to make a recommendation for or against the prophylactic use of vancomycin compared with other antimicrobial agents to reduce infections with MRSA. Grade of recommendation: I (insufficient evidence)	The effectiveness of vancomycin compared to other antimicrobial agents in reducing MRSA infections in spinal surgery is not clear. Some studies suggest that vancomycin may be effective in preventing MRSA infections in spinal surgery for patients with a history of MRSA infection, but others have found no significant difference between vancomycin and other antibiotics.	The study from PubMed suggests that a quality improvement protocol that includes MRSA screening and decolonization in patients undergoing lumbar fusion and/or decompression surgery significantly reduces the rate of postoperative infections, including MRSA and MSSA infections. Significant predictors of infection post-surgery were younger age, male sex, diabetic status, higher American Society of Anesthesiologists physical status classification grades, and undergoing spinal fusion. For a comprehensive understanding, further research from multiple sources would be necessary to compare the effectiveness of vancomycin to other antimicrobial agents specifically in spine surgery settings.
Complications	What are the incidence and severity of complications/adverse events resulting from the use of prophylactic antibiotics?	Consensus statement: Reported isolated complications related to prophylactic antibiotics include flushing, hypotension, rashes, intramembranous colitis and, most seriously, Stevens-Johnson syndrome.	The incidence and severity of complications/adverse events resulting from the use of prophylactic antibiotics may include allergic reactions, antibiotic resistance, superinfections, nausea, vomiting, diarrhea, other gastrointestinal symptoms, headache, dizziness, and other side effects.	The use of prophylactic antibiotics can lead to complications such as higher incidence of nosocomial pneumonia, especially with prolonged use, and the development of resistant or gram-negative bacteria. The benefits of antimicrobial prophylaxis in surgery are mostly limited to the first 24 hours postoperatively, with potential harms like surgical site infection, acute kidney injury, and <i>C. difficile</i> infection associated with longer durations of use. Continuous prophylaxis can reduce urinary tract infection frequency but raises concerns about increased bacterial resistance. Serious complications, like anaphylaxis, are rare but can occur, as with cephalosporins which have a reported incidence of one per thousand. In general, prophylactic antibiotics are effective in reducing the incidence of certain infections, but careful consideration is needed due to the risk of severe adverse events and antibiotic resistance.

(Continued)

Table 2. A comparison of the clinical questions and North American Spine Society (NASS) guidelines for antibiotics prophylaxis in spine surgery with corresponding ChatGPT model recommendations (GPT-3.5 & GPT-4.0) (Continued)

Category	NASS question	NASS guidelines	GPT-3.5 recommendation	GPT-4.0 recommendation
Complications	What strategies can be implemented to minimize complications/adverse events resulting from the use of prophylactic antibiotics in spine surgery?	<p>Consensus statement: In typical, uncomplicated spinal procedures, a single dose of preoperative prophylactic antibiotics with intraoperative redosing as needed is suggested to reduce the risk of complications/adverse events.</p> <p>Reported isolated complications/adverse events related to prophylactic antibiotics are discussed in the previous section and include: flushing, hypotension, rashes, intramembranous colitis and, most seriously, Stevens-Johnson syndrome.</p>	<p>The use of prophylactic antibiotics in spine surgery carries the risk of adverse events, such as allergic reactions and antibiotic resistance. Some strategies that can minimize adverse events include selecting antibiotics based on patient risk factors, proper timing of administration, limiting the duration of prophylaxis to the perioperative period, monitoring the patient's response, considering relevant patient factors, allergy testing, and educating medical practitioners on appropriate use of prophylactic antibiotics.</p>	<p>Minimizing complications from prophylactic antibiotics in spine surgery involves a few evidence-based strategies: Duration of Antibiotics: A study indicated that there was no significant difference in the rate of SSIs between patients who received 24 hours of postoperative antibiotics and those who did not. This suggests that extending prophylactic antibiotic use beyond the immediate preoperative period may not be necessary and could avoid potential risks like antibiotic-resistant infections and <i>C. difficile</i> infections. Evidence-Based Guidelines: The NASS clinical guideline for the use of prophylactic antibiotics in spine surgery, which is based on high-quality clinical literature, provides recommendations on antibiotic protocols, dosing, and special considerations related to comorbidities to prevent SSIs. These strategies are part of a broader approach that includes optimizing surgical techniques, maintaining sterile environments, and utilizing targeted antibiotics based on the surgery type and patient-specific factors. The complete set of recommendations and evidence can be found in the full guideline documents provided by NASS on their website.</p>

ChatGPT, chat generative pre-trained transformer.

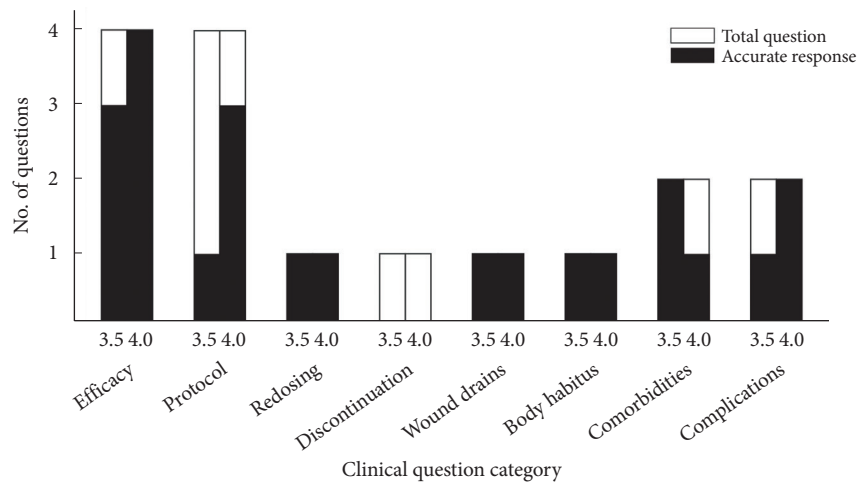


Fig. 3. Breakdown in GPT-3.5 and GPT-4.0 model performance by subtopic. GPT, generative pre-trained transformer.

but there was insufficient evidence to definitively state whether obesity increased the risk of SSIs. GPT-3.5 correctly identified the relationship between SSI risk and diabetes but not for obesity,¹⁵ overconfidently stating that obesity increases risk for SSI despite insufficient evidence to support this point. While some recent studies have reported an impact of obesity on SSIs,^{16,17} others have suggested that comorbidities often seen in obese patients, such as diabetes, may be the ultimate driver of increased SSI rates,¹⁸ while others suggest that body habitus, rather than obesity, affects SSI rate.¹⁹ In comparison, GPT-4.0 correctly answered this question and refrained from making incorrect assumptions by conveying the ambiguity of this topic.

2. Protocol

Three of 4 protocol questions concerned the recommended drugs, dosages, administration routes, and timing for general spine surgery and spine surgery with or without implants. The clinical guidelines state that the superiority of any drug, dose, or route of administration has not been demonstrated, and since the release of these guidelines, the literature has yet to determine a superior protocol for general spine surgery.^{20,21} While several studies have investigated varying prophylaxis protocols, the evidence remains contradictory and insufficient to merit a recommendation.²⁰ In response to these questions, GPT-3.5 failed to definitively state that no single drug, dosage, or administration route was recommended while GPT-4.0 accurately answered all three. For 2 questions, GPT-3.5 explicitly stated that cefazolin was the recommended first-line antibiotic, but this is inconclusive in the literature.

The last question asked for a reasonable algorithmic approach to antibiotic selection for a given patient. The clinical guidelines

recommend one single preoperative dose of the antibiotic of choice, with intraoperative redosing as needed, for simple and uncomplicated spinal procedures with consideration of additional gram-negative coverage and/or the application of intra-wound vancomycin or gentamicin for complicated cases. Both ChatGPT models gave a general list of steps that could be taken to choose appropriate antibiotics and protocols rather than the correct recommendations outlined for simple and complex procedures.

3. Redosing and Discontinuation

One question asked about intraoperative redosing and drug recommendations. To the best of our knowledge, there have been no studies directly examining the effects of redosing on SSI incidence, but many studies that have included intraoperative redosing protocols in their cohorts agree with a standard redosing protocol within 3 to 4 hours of the initial dose,^{22,23} in line with NASS guidelines. The guidelines also state that the superiority of one drug has not been definitively established. ChatGPT models gave individual redosing recommendations for several common antibiotics and recommended redosing after 3–4 hours for cefazolin, which we rated as accurate.

Another question asked whether discontinuation of antibiotic prophylaxis after 24 hours had an effect on SSI incidence compared to a longer duration. ChatGPT correctly answered that discontinuation at 24 hours does not increase SSI risk compared to longer durations, which has been reported by several studies.^{13,20,24,25} While ChatGPT models are not accurate compared to the outdated guidelines from 2013, they are both accurate in their answers.

4. Wound Drains and Body Habitus

One question asked if discontinuation of antibiotic prophylaxis at 24 hours affected incidence of SSIs compared to discontinuation at the time of wound drain removal. At the time of the release of the NASS guidelines, few studies had examined the effect of prophylaxis discontinuation with wound drains, resulting in insufficient evidence to make a recommendation.^{26,27} Since the release of the guidelines, ambiguity remains in the literature, although one study has shown that discontinuing antibiotics at the time of drain removal does not decrease the rate of SSIs.²⁸ ChatGPT correctly stated that no recommendation could be made.

Another question asked if the recommended prophylaxis protocol differs based on body habitus or body mass index. While the literature has shown that obese patients may face higher risks of SSIs following posterior cervical spine surgery²⁹ and lumbar spine surgery,³⁰ there is currently no consensus on alternative antibiotic prophylaxis protocols for obese patients. ChatGPT adequately conveyed that the evidence is too limited to suggest an altered antibiotic regimen based on body habitus.

5. Comorbidities

Two questions addressed comorbidities associated with antibiotic prophylaxis. One asked if comorbidities other than obesity necessitate altered antibiotic prophylaxis protocols. According to the NASS guidelines, surgeons should consider altered protocols for patients with various comorbidities; for example, diabetic patients may benefit from intraoperative redosing,³¹ and the intrawound application of vancomycin has been reported to significantly decrease the incidence of deep and superficial SSIs in patients with comorbidities.³² ChatGPT adequately responded that there are several alternative protocols that have been shown to help prevent SSIs in patients with various comorbidities, with GPT-4.0 directly citing the NASS guideline.

A second question asked if vancomycin reduces infections by methicillin-resistant *Staphylococcus aureus* (MRSA) in patients with a history of MRSA infection. At the time of the release of the guidelines, there was insufficient evidence to make a recommendation. While a recent meta-analysis has suggested that local application of vancomycin powder can reduce the incidence of MRSA infections following posterior spine surgery³³ and elective lumbar procedures,³⁴ other studies have reported no impact of vancomycin on SSI or gram-negative infection incidence.^{35,36} Therefore, the evidence remains insufficient to make a recommendation. GPT-3.5 was correct in that it did not give

a recommendation for an approach and stated the use of vancomycin for reducing MRSA infection should be based on patient risk factors, the local prevalence of MRSA, and institutional guidelines. Meanwhile, GPT-4.0 gave one study that suggested a quality improvement approach would decrease postoperative infections including MRSA and MSSA infections without reporting insufficient evidence, hence making this answer incorrect.

6. Complications

One question concerned the general incidence and severity of complications arising from the use of antibiotic prophylaxis in spine surgery. The reported incidence of antibiotics-related complications is low,^{12,37} although flushing, hypotension,³⁸ and rash³⁹ have been reported in rare cases. Both ChatGPT models answered the question by listing several potential complications that were not directly listed in the guidelines.

A second question asked what strategies can be implemented to reduce the risk of complications related to the use of antibiotic prophylaxis. The NASS guidelines recommend a single dose of preoperative prophylactic antibiotics with intraoperative redosing as needed for uncomplicated spine procedures. GPT-3.5 provided a long list of general strategies, some of which were accurate and clinically relevant but others were incomplete, overgeneralized and contradictory. For example, it recommended that the duration of prophylaxis be limited to the perioperative period, stating prolonged prophylaxis is known to result in increased risk of adverse events and infections; however, in the discontinuation category, GPT-3.5 reported no difference in infections between perioperative and prolonged durations of prophylaxis, and in the comorbidities category, it suggested that prolonged duration of prophylaxis could help reduce SSI risk in patients with comorbidities. GPT-4.0 correctly answered this question by citing the NASS guidelines and pointing to its recommendations.

7. General Discussion

ChatGPT demonstrated an impressive ability to accurately answer clinical questions given that it was not specifically trained on a medical dataset. While both models showed good comprehension of the prompts, GPT-3.5 model's performance was limited by its tendency to give overly confident responses for prompts with non-conclusive evidence and its inability to identify the most significant elements in its response. For example, in the protocol category, GPT-3.5 did not definitively state that there was not enough evidence to recommend a specific proto-

col, instead defaulting to a general statement that several factors should be considered. However, in other cases, it was able to convey lack of consensus on the given question in its response—for example, in the body habitus and comorbidities category. GPT-4.0 model's responses not only had higher accuracy but also cited the NASS guideline as direct evidence in 10 of the 16 questions it correctly answered. This points to its exceptional ability to extract the most relevant research available compared to GPT-3.5 in its recommendation for antibiotic prophylaxis in spine surgery.

It is not surprising that GPT-4.0 outperformed GPT-3.5. The newer model was trained on a larger and more recent dataset than its predecessor, although the details of this data set and training process have not been released to the public. Additionally, human reviewers and AI were used to provide reinforcement learning for GPT-4.0. These improvements have been shown to decrease GPT-4.0's tendency to "hallucinate" responses compared to its predecessor and give up to 40% more factual information.¹

One of the most significant limitations to the application of ChatGPT in clinical settings is the unpredictability of the model's responses. ChatGPT is programmed to give plausible-sounding responses, but the actual information contained in the response may be incorrect, misleading, or completely fabricated, especially in the widely available and free GPT-3.5 model. We found GPT-3.5 was overconfident in 4 of 16 questions (25%), sometimes fabricating studies and citing nonexistent papers to prove its point. However, while several responses containing fabricated references were ultimately correct, the information could not be easily confirmed. A significant obstacle facing ChatGPT and other LLMs is their inclination to provide conclusive responses even if it would be more accurate to indicate a lack of agreement. This has been labeled artificial hallucination, where LLMs tend to produce plausible statements without any real basis in their training data.⁴⁰

GPT-4.0 is a more enhanced model that ameliorates this effect by including direct citations to its sources, with 10 of 16 questions (62.5%) ultimately citing the NASS guideline to back up its correct answers. However, another obstacle facing ChatGPT in clinical settings that has been noted by other researchers is its inconsistency. Howard et al.⁹ reported that the clinical recommendations provided by ChatGPT often changed on repeated questioning. While we did not prompt ChatGPT the same questions multiple times, we did observe minor inconsistencies in its responses, namely its shifting recommendations for prolonged duration of prophylaxis.

GPT should be used with proper care and caution in clinical settings. GPT-4.0's ability to cite literature has the potential to be used by clinicians for consolidating disparate sources of information and summarizing them. Even still, it is not always accurate and should be further validated when being used to guide important clinical decision-making.

8. Future Directions

Development of next generation LLMs is already underway with the future release of GPT-4.0 Turbo,¹ which in addition to providing more complex responses to textual prompts allows for pretraining and domain specific models such as those trained on medical literature. Other medical domain specific models such as Med-PaLM are in development, but do not have public access yet. The excitement around these models and their potential widespread applications must be counterbalanced by sufficient research investigating the limitations and potential hazards of applying this technology in clinical settings. The application of LLMs in complex clinical settings can present serious ethical issues for both patient care and concerning liability. There is the possibility for LLMs to give inaccurate responses, particularly when prompted for medical advice given the imbalance of medical information in the training data. These responses must be aligned with clinical knowledge. In cases where a discrepancy exists and patients are affected, the question of liability and minimizing patient harm is of ethical importance. Future research should continue to assess the performance of LLMs, including the latest version of ChatGPT, in order to provide medical professionals with evidence-based guidelines for the use of LLMs. Furthermore, these assessments could benefit from deeper investigations into prompt engineering, which is the process of structuring prompts in order to optimize the comprehension of the LLM. Small changes in the structure of prompts could lead to potentially major discrepancies in responses, which present additional challenges to the application of LLMs in clinical settings that must be addressed. Real-world clinical applications of these models include responding to common patient questions and automated analysis of medical records. However, there are still challenges such as ensuring proper data privacy and accuracy of responses.

9. Limitations

A limitation of this study is that the NASS Evidence-Based Clinical Guidelines for Multidisciplinary Spinal Care have not been updated since 2013 and may now be outdated, while ChatGPT has been trained on textual data from the Internet up to

2021. However, this source represents the most recent organizational guidelines for spinal prophylaxis. We also compared ChatGPT's responses to the most up-to-date literature, with GPT-3.5 trained with data up to 2021 and GPT-4.0 to April 2023,¹ therefore, more recent literature may not be entirely reflected in its responses. Another limitation of this study is that LLMs may not always respond identically to a given prompt, and are often updated making comparisons between responses more difficult. Moreover, this study is limited by the prompts that were used as they were verbatim from the guidelines and subject to the biases inherent in their structuring and phrasing. Given the nature of AI, advances happen rapidly and newer models are currently being developed which may perform better or be trained on more clinically focused data. Finally, the training data used on the models has never been publicly stated, and as such there may be biases and gaps in medical knowledge that are not publicly known.

CONCLUSION

This study suggests that we continue to approach LLMs with caution and optimism. While GPT-3.5 was able to generate clinically relevant antibiotic use guidelines for spinal surgery, its inability to identify the most important aspects of the comparable guidelines and its redundancy, fabrication of citations, and inconsistency are obstacles to the application of the model in clinical settings. At the same time, GPT-4.0 was almost 20% better in its response accuracy and was able to cite the NASS guideline 62.5% of the time. While ChatGPT has far-reaching potential in directing users to evidence-based research as a basis of its findings, scrutiny should still be exercised in using ChatGPT reliably at this time. Medical practitioners should be cautioned against relying on ChatGPT for clinical recommendations on antibiotic prophylaxis in spine surgery without close scrutiny of the medical literature.

NOTES

Conflict of Interest: Jun S. Kim: Stryker, Paid consultant. Samuel Kang-Wook Cho, FAAOS: AAOS, Board or committee member; American Orthopaedic Association, Board or committee member; AOSpine North America, Board or committee member; Cervical Spine Research Society, Board or committee member; Globus Medical, IP royalties; North American Spine Society, Board or committee member; Scoliosis Research Society, Board or committee member; Stryker, Paid consultant. The

following authors have nothing to disclose: Bashar Zaidat, Nancy Shrestha, Wasil Ahmed, Rami Rajjoub, Ashley M. Rosenberg, Timothy Hoang, Mateo Restrepo Mejia, Akiro H. Duey, Justin E. Tang.

Funding/Support: This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution: Conceptualization: BZ, SKC; Formal Analysis: BZ; Investigation: BZ; Methodology: BZ; Project Administration: BZ, JSK, SKC; Writing - Original Draft: BZ, NS, AMR; Writing - Review & Editing: BZ, NS, AMR, WA, RR, TH, MRM, AHD, JET, JSK, SKC.

ORCID

Bashar Zaidat: 0000-0002-8823-720X
 Nancy Shrestha: 0000-0003-0314-7748
 Ashley M. Rosenberg: 0000-0001-9029-445X
 Wasil Ahmed: 0009-0001-0904-1891
 Rami Rajjoub: 0009-0005-2990-7874
 Timothy Hoang: 0000-0002-3012-3461
 Mateo Restrepo Mejia: 0009-0003-0457-3308
 Akiro H. Duey: 0000-0003-0308-0512
 Justin E. Tang: 0000-0003-4544-3262
 Jun S. Kim: 0000-0002-6114-2673
 Samuel K. Cho: 0000-0001-7511-2486

REFERENCES

1. OpenAi. GPT-4 Technical Report. arXiv:2303.08774v4 [Preprint]. 2023. Available from: <https://arxiv.org/abs/2303.08774>.
2. Zaidat B, Tang J, Arvind V, et al. Can a novel natural language processing model and artificial intelligence automatically generate billing codes from spine surgical operative notes? *Global Spine J* 2023 Mar 18:21925682231164935. doi: 10.1177/21925682231164935. [Epub].
3. Kim JS, Vivas A, Arvind V, et al. Can natural language processing and artificial intelligence automate the generation of billing codes from operative note dictations? *Global Spine J* 2023;13:1946-55.
4. Marafino BJ, Park M, Davies JM, et al. Validation of prediction models for critical care outcomes using natural language processing of electronic health record data. *JAMA Network Open* 2018;1:e185097.
5. Zhang X, Kim J, Patzer RE, et al. Prediction of emergency department hospital admission based on natural language processing and neural networks. *Methods Inf Med* 2017;56:

- 377-89.
6. Lin SY, Mahoney MR, Sinsky CA. Ten ways artificial intelligence will transform primary care. *J Gen Intern Med* 2019; 34:1626-30.
 7. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS Digital Health* 2023;2:e0000198.
 8. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the united states medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ* 2023;9: e45312.
 9. Howard A, Hope W, Gerada A. ChatGPT and antimicrobial advice: the end of the consulting infection doctor? *Lancet Infect Dis* 2023;23:405-6.
 10. Shaffer WO, Baisden JL, Fernand R, et al. An evidence-based clinical guideline for antibiotic prophylaxis in spine surgery. *Spine J* 2013;13:1387-92.
 11. Amelot A, Riche M, Latreille S, et al. Antimicrobial prophylaxis in noninstrumented spine surgery: a prospective study to determine efficacy and drawbacks. *J Neurosurg Spine* 2021;35:366-75.
 12. Petignat C, Francioli P, Harbarth S, et al. Cefuroxime prophylaxis is effective in noninstrumented spine surgery: a double-blind, placebo-controlled study. *Spine (Phila Pa 1976)* 2008;33:1919-24.
 13. Shawky Abdelgawaad A, El Sadik MHM, Hassan KM, et al. Perioperative antibiotic prophylaxis in spinal surgery. *SICOT J* 2021;7:31.
 14. Zhou J, Wang R, Huo X, et al. Incidence of surgical site infection after spine surgery: a systematic review and meta-analysis. *Spine (Phila Pa 1976)* 2020;45:208-16.
 15. Chen S, Anderson MV, Cheng WK, et al. Diabetes associated with increased surgical site infections in spinal arthrodesis. *Clin Orthop Relat Res* 2009;467:1670-3.
 16. Lim S, Edelstein AI, Patel AA, et al. Risk factors for postoperative infections after single-level lumbar fusion surgery. *Spine (Phila Pa 1976)* 2018;43:215-22.
 17. Jiang J, Teng Y, Fan Z, et al. Does obesity affect the surgical outcome and complication rates of spinal surgery? a meta-analysis. *Clin Orthop Relat Res* 2014;472:968-75.
 18. Seicean A, Alan N, Seicean S, et al. Impact of Increased body mass index on outcomes of elective spinal surgery. *Spine (Phila Pa 1976)* 2014;39:1520-30.
 19. Mehta AI, Babu R, Karikari IO, et al. 2012 Young investigator award winner: the distribution of body mass as a significant risk factor for lumbar spinal fusion postoperative infections. *Spine (Phila Pa 1976)* 2012;37:1652-6.
 20. Yao R, Tan T, Tee JW, et al. Prophylaxis of surgical site infection in adult spine surgery: a systematic review. *J Clin Neurosci* 2018;52:5-25.
 21. Nishant, Kailash KK, Vijayraghavan PV. Prospective randomized study for antibiotic prophylaxis in spine surgery: choice of drug, dosage, and timing. *Asian Spine J* 2013;7:196-203.
 22. Hellbusch LC, Helzer-Julien M, Doran SE, et al. Single-dose vs multiple-dose antibiotic prophylaxis in instrumented lumbar fusion--a prospective study. *Surg Neurol* 2008;70: 622-7; discussion 627.
 23. Takahashi H, Wada A, Iida Y, et al. Antimicrobial prophylaxis for spinal surgery. *J Orthop Sci* 2009;14:40-4.
 24. Marimuthu C, Abraham VT, Ravichandran M, et al. Antimicrobial prophylaxis in instrumented spinal fusion surgery: a comparative analysis of 24-hour and 72-hour dosages. *Asian Spine J* 2016;10:1018-22.
 25. Kim B, Moon SH, Moon ES, et al. Antibiotic prophylaxis for spinal surgery: comparison between 48 and 72-hour AMP protocols. *Asian Spine J* 2010;4:71-6.
 26. Pavel A, Smith RL, Ballard A, et al. Prophylactic antibiotics in elective orthopedic surgery: a prospective study of 1,591 cases. *South Med J* 1977;70 Suppl 1:50-5.
 27. Kanayama M, Hashimoto T, Shigenobu K, et al. Effective prevention of surgical site infection using a Centers for Disease Control and Prevention guideline-based antimicrobial prophylaxis in lumbar spine surgery. *J Neurosurg Spine* 2007;6:327-9.
 28. Takemoto RC, Lonner B, Andres T, et al. Appropriateness of twenty-four-hour antibiotic prophylaxis after spinal surgery in which a drain is utilized: a prospective randomized study. *J Bone Joint Surg Am* 2015;97:979-86.
 29. Sebastian A, Huddleston P 3rd, Kakar S, et al. Risk factors for surgical site infection after posterior cervical spine surgery: an analysis of 5,441 patients from the ACS NSQIP 2005-2012. *Spine J* 2016;16:504-9.
 30. Meng F, Cao J, Meng X. Risk factors for surgical site infections following spinal surgery. *J Clin Neurosci* 2015;22:1862-6.
 31. Zhang X, Li T, Li Y, et al. Protective effect of intraoperative re-dose of prophylactic antibiotics on surgical site infection in diabetic patients: a retrospective cohort study. *Ann Transl Med* 2019;7:96.
 32. Lemans JVC, Öner FC, Wijdicks SPJ, et al. The efficacy of intrawound vancomycin powder and povidone-iodine irri-

- gation to prevent surgical site infections in complex instrumented spine surgery. *Spine J* 2019;19:1648-56.
33. Luo H, Ren Y, Su Y, et al. Intraoperative vancomycin powder to reduce surgical site infections after posterior spine surgery: a systematic review and meta-analysis. *EFORT Open Rev* 2022;7:109-21.
 34. Conaway W, Lambrechts MJ, D'Antonio ND, et al. MRSA Prophylaxis in spine surgery decreases postoperative infections. *Clin Spine Surg* 2023;36:E153-9.
 35. Xiong GX, Greene NE, Hershman SH, et al. Nasal screening for methicillin-resistant *Staphylococcus aureus* does not reduce surgical site infection after primary lumbar fusion. *Spine J* 2022;22:113-25.
 36. Gande A, Rosinski A, Cunningham T, et al. Selection pressures of vancomycin powder use in spine surgery: a meta-analysis. *Spine J* 2019;19:1076-84.
 37. Sweet FA, Roh M, Sliva C. Intrawound application of vancomycin for prophylaxis in instrumented thoracolumbar fusions: efficacy, drug levels, and patient outcomes. *Spine (Phila Pa 1976)* 2011;36:2084-8.
 38. Pons VG, Denlinger SL, Guglielmo BJ, et al. Ceftizoxime versus vancomycin and gentamicin in neurosurgical prophylaxis: a randomized, prospective, blinded clinical study. *Neurosurgery* 1993;33:416-22; discussion 422-3.
 39. Mastronardi L, Rychlicki F, Tatta C, et al. Spondylodiscitis after lumbar microdiscectomy: effectiveness of two protocols of intraoperative antibiotic prophylaxis in 1167 cases. *Neurosurg Rev* 2005;28:303-7.
 40. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus* 2023;15:e35179.