Department of Orthopaedic Surgery Faculty Papers

Department of Orthopaedic Surgery

9-4-2024

Assessing the Accuracy and Utility of ChatGPT Responses to Patient Questions Regarding Posterior Lumbar Decompression

Alec Giakas

Rajkishen Narayanan

Teeto Ezeonu

Jonathan Dalton

Yunsoo Lee

See next page for additional authors

Authors

Alec Giakas, Rajkishen Narayanan, Teeto Ezeonu, Jonathan Dalton, Yunsoo Lee, Tyler Henry, John Mangan, Gregory Schroeder, Alex Vaccaro, and Christopher Kepler

Original Article

Assessing the accuracy and utility of ChatGPT responses to patient questions regarding posterior lumbar decompression

Alec M. Giakas (ID)**, Rajkishen Narayanan, Teeto Ezeonu, Jonathan Dalton, Yunsoo Lee, Tyler Henry, John Mangan, Gregory Schroeder, Alexander Vaccaro, Christopher Kepler**

Department of Orthopaedic Surgery, Rothman Institute at Thomas Jefferson University, Philadelphia, PA 19107, USA.

Correspondence to: Dr. Rajkishen Narayanan, Department of Orthopaedic Surgery, Rothman Institute at Thomas Jefferson University, 925 Chestnut Street Floor 5, Philadelphia, PA 19107, USA. E-mail: rajkishen.narayanan@gmail.com

Abstract

Aim: To examine the clinical accuracy and applicability of ChatGPT answers to commonly asked questions from patients considering posterior lumbar decompression (PLD).

Methods: A literature review was conducted to identify 10 questions that encompass some of the most common questions and concerns patients may have regarding lumbar decompression surgery. The selected questions were then posed to ChatGPT. Initial responses were then recorded, and no follow-up or clarifying questions were permitted. Two attending fellowship-trained spine surgeons then graded each response from the chatbot using a modified Global Quality Scale to evaluate ChatGPT's accuracy and utility. The surgeons then analyzed each question, providing evidence-based justifications for the scores.

Results: Minimum scores across all ten questions would lead to a total score of 20, whereas a maximum score would be 100. ChatGPT's responses in this analysis earned a score of 59, just under an average score of 3, when evaluated by two attending spine surgeons. A score of 3 denoted a somewhat useful response of moderate quality, with some important information adequately discussed but some poorly discussed.

Conclusion: ChatGPT has the ability to provide broadly useful responses to common preoperative questions that patients may have when considering undergoing PLD. ChatGPT has excellent utility in providing background

information to patients and in helping them become more informed about their pathology in general. However, it often lacks the specific patient context necessary to provide patients with personalized, accurate insights into their prognosis and medical options.

INTRODUCTION

In recent decades, internet search engines have transformed the way patients seek health information and interact with the health care system[1,2]. Accordingly, almost 75% of adult internet users have searched for health information online. Within the orthopedic patient population, up to two-thirds of patients have used the Internet to search for information about their conditions[3,4]. Recently, open-access artificial intelligence (AI) has become a popular tool for information-gathering in healthcare[5-7]. ChatGPT, launched in November 2022, is an AI-powered language processing tool that has the ability to integrate billions of datapoints and fine-tune successive responses to mimic human conversation[8,9]. Recently, the AI chatbot garnered attention within the medical community for earning a passing grade on a United States Medical Licensing Examination practice test, an exam that involves the application of knowledge and critical thinking[10].

As patients become more familiar with AI platforms, they may increasingly look to these programs to provide instantaneous answers to questions about their health. Previous studies have examined ChatGPT's utility in clinical decision making and patient education within orthopedics[11-16]. In a recent paper, Mika *et al.* attempted to determine if ChatGPT could answer frequently asked questions related to total hip arthroplasty[17]. In their conclusion, the authors noted ChatGPT's utility in the clinical space by providing evidence-based responses, but acknowledged that responses often required at least some clarification. Similarly, in another paper, AI-generated answers to common patient questions regarding minimally invasive spine surgery (MISS) were explored. While authors acknowledged ChatGPT's ability to accurately describe procedures, indications, outcomes, and complications of MISS, they also noted its tendency to make unsubstantiated claims and potentially confuse patients as they make decisions regarding their treatment[18]. Given the concerning discrepancies that these prior studies have identified, additional scrutiny into AI's ability to accurately aid patients and physicians in the shared decision-making process is necessary[19].

Degenerative spine conditions are a prominent health concern, particularly within the aging population, that can cause patients significant physical, mental, and emotional distress. Posterior lumbar decompression (PLD), a procedure performed to treat degenerative spine conditions, is one of the most commonly performed procedures by spine surgeons[20,21]. The utility and limitations of ChatGPT in answering common preoperative questions for patients considering PLD have yet to be explored. Furthermore, while prior studies in the adult reconstruction literature have attempted to score and analyze ChatGPT responses, no previous study has methodologically scored AI's responses to common questions asked by patients considering spine surgery. As such, the purpose of the present study is to evaluate the clinical accuracy and applicability of ChatGPT answers to commonly asked questions from patients considering PLD.

METHODS

Methods

Question selection

First, a literature review was performed for commonly asked questions by patients considering orthopedic surgery - 17 questions were identified[4,12,15,18]. Next, we reviewed the "frequently asked questions" on the websites of 10 prominent healthcare institutions to further identify potential patient questions. 5 additional questions were pooled from this website review. Lastly, we queried ChatGPT itself for common questions patients ask when considering spine surgery. 12 questions were elicited from ChatGPT itself. Ultimately, 10 questions were selected from these three sources by two attending, fellowship-trained orthopedic spine surgeons at a high-volume academic institution. The questions selected were felt to encompass some of the most common questions and pressing concerns patients may have regarding lumbar decompression surgery.

ChatGPT queries

The selected 10 questions were then posed to ChatGPT (Version 4.0) using its free online interface (https://chat.openai.com). Initial responses were then recorded, and no follow-up or clarifying questions were permitted. Responses have been truncated in this manuscript for the purpose of publication. Complete responses can be found in the Supplementary Materials.

Analysis

The two previously mentioned attending orthopedic spine surgeons analyzed each response from the chatbot using a modified Global Quality Scale to assign a score out of 5 possible points evaluating ChatGPT's accuracy and utility [Table 1][22]. A score of 5 was reserved for evidence-based, comprehensive responses of excellent quality that were both accurate and very useful for patients. A score of 4 denoted good quality, useful responses with minimal unsubstantiated claims. A score of 3 signified a response of moderate quality, somewhat useful for patients with some poorly discussed important information. A score of 2 denoted a generally poor-quality response, with many important topics missing and some unsubstantiated claims that have very little use to patients. Finally, a score of 1 signified a poor-quality response lacking any utility with missing information and multiple unsubstantiated claims. Scores given to all responses by each attending were recorded [Table 1].

RESULTS

1. When should I get lumbar decompression surgery?

The decision to undergo lumbar decompression surgery should be made in consultation with a qualified healthcare professional, preferably a spine specialist or orthopedic surgeon. The timing of surgery depends on various factors, including the severity of your symptoms, the underlying condition causing the symptoms, your overall health, and the effectiveness of non-surgical treatments. Here are some considerations for determining when you might consider lumbar decompression surgery:

(1) Conservative treatment failure
(2) Persistent severe pain
(3) Progressive neurological deficits
(4) Loss of function
(5) Quality of life impact
(6) Emergent cases
(7) Health status
(8) Patient preference

Table 1. Modified Global Quality Scale items

#	Item
1	Poor quality, multiple unsubstantiated claims, most information missing. Not at all useful for patients
2	Generally poor quality, some unsubstantiated claims, some information listed but many important topics missing. Very limited use to patients
3	Moderate quality, some important information is adequately discussed, but others poorly discussed. Somewhat useful for patients
4	Good quality, minimal unsubstantiated claims, most of the relevant information is listed, but some topics not covered. Useful for patients
5	Excellent quality, comprehensive claims that are accurate and overall evidence-based. Very useful for patients

Attending 1 Score: 4
Attending 2 Score: 3

Analysis: ChatGPT provided a basic, but diverse general overview of why a patient might consider PLD. The answer correctly identifies several general reasons for considering PLD, notably including "emergent cases". However, it does not mention any specific diagnoses that would be indications for surgery, such as neurogenic claudication or radiculopathy[23-25]. Responses such as "health status" and "patient preference" are quite vague. In particular, "patient preference" may be interpreted broadly as indicating that surgery is simply optional, and many patients may use this as justification to avoid surgery. This answer should be qualified with a warning that this is highly dependent on both the clinical scenario and underlying pathology and that irreversible pain or neurologic deficit may result from delaying or avoiding surgery without consulting with a spine surgeon. Finally, this answer overall lacks evidence-based responses that might be useful to patients considering surgery.

2. What will happen if I delay lumbar decompression surgery?
Delaying lumbar decompression surgery might have various implications depending on your specific condition and symptoms. While some conditions may tolerate a certain degree of delay, others might worsen over time. It is important to consult with a healthcare professional, ideally a spine specialist, to understand the potential consequences of delaying surgery in your individual case. Here are some general considerations:

(1) Progression of symptoms: If you are experiencing pain, weakness, numbness, or other neurological symptoms due to nerve compression, delaying surgery might lead to the progression of these symptoms. Nerve compression can cause permanent damage if left untreated.
(2) Functional limitations: Nerve compression can affect your ability to perform daily activities, work, and participate in recreational activities. Delaying surgery might limit your quality of life and functional abilities.
(3) Quality of life: The impact of your symptoms on your overall well-being and quality of life is an important factor to consider. If your symptoms are significantly affecting your day-to-day life, addressing them sooner through surgery might be beneficial.
(4) Impact on surgical outcome: In some cases, delaying surgery might lead to more complex surgery later, as conditions can progress, and tissue changes might occur over time.
(5) Psychological impact: Chronic pain and limited mobility can have psychological effects, including anxiety and depression. Addressing these issues through timely treatment might prevent their exacerbation.
(6) Risk of complications: If your symptoms are related to structural issues such as herniated discs, delaying surgery could increase the risk of complications associated with nerve compression or other related problems.

Attending 1 Score: 3
Attending 2 Score: 3

Analysis: This ChatGPT response provides a somewhat useful overview of the potential results of delaying surgery. However, it lacks any specific information as to the consequences of "nerve damage" that might be useful to the patient considering surgery. Instead of providing specific information, such as increasing muscle weakness, or loss of bowel or bladder function, the response focuses more on how delaying surgery may impact a patient's quality of life in general[26,27]. This response is reasonable, but lacks the specificity and comprehensiveness needed by patients considering PLD. The complications associated with delaying lumbar decompression surgery are vague and do not convey the potential severity involved with this decision. Additionally, no data are provided to the patient in terms of the percentage of complications from non-operative treatment, which may provide some context to the patient trying to understand the scope of this choice.

3. What are the alternatives to lumbar decompression surgery?
There are several alternatives to lumbar decompression surgery that may be considered depending on your specific condition, symptoms, and overall health. It is important to discuss these alternatives with a qualified healthcare professional, preferably a spine specialist, to determine the most suitable treatment plan for your individual situation. Here are some non-surgical alternatives to lumbar decompression surgery:

(1) Physical therapy: Physical therapy can help strengthen muscles, improve flexibility, and provide strategies to manage pain. Therapists can design exercise programs tailored to your needs.
(2) Medications: Non-prescription pain relievers (e.g., acetaminophen, NSAIDs) or prescription medications (e.g., muscle relaxants, neuropathic pain medications) can help manage pain and inflammation.
(3) Epidural steroid injections: These injections can help reduce inflammation and provide temporary pain relief. They are often used when pain is primarily due to nerve irritation.
(4) Lifestyle modifications: Making changes to your daily habits, such as maintaining a healthy weight, improving posture, and avoiding activities that exacerbate your symptoms, can help manage pain and prevent worsening of your condition.
(5) Chiropractic care: Chiropractors use manual manipulation techniques to improve spinal alignment and alleviate pain. However, caution should be exercised if chiropractic care is considered for certain conditions.
(6) Bracing: Some conditions might benefit from wearing a supportive brace to help stabilize the spine and reduce pain during certain activities.
(7) Watchful waiting: In some cases, especially for mild symptoms, your healthcare provider might recommend monitoring your condition over time to see if it improves or worsens before considering more invasive treatments.

Attending 1 Score: 3
Attending 2 Score: 4

Analysis: This response provides a good summation of alternative options to surgery that might be useful to patients. However, it gives minimal detail regarding when these alternatives might be sufficient treatments, and thus risks patients delaying or avoiding necessary surgery. In such instances when surgery is indicated, initiating or continuing non-operative treatment might lead to progressive and permanent neurological impairment for patients[28,29]. As such, while the response lists, in appropriate detail, options that might be appropriate for some patients, pursuing these options might be highly detrimental for other patients.

Furthermore, these responses do not elaborate on the potential adverse effects associated with various treatment options, thus providing a weak and incomplete risk-benefit analysis for patients to consider when deciding whether to proceed with PLD. Within the descriptions of the alternative therapies, even for the appropriate patient, there is minimal detail on what these methods entail. For example, the physical therapy response does not describe what muscles are strengthened or why, nor how much time may elapse prior to symptom alleviation.

4. What are the risks of lumbar decompression surgery?

Lumbar decompression surgery is generally considered safe and effective for relieving symptoms caused by nerve compression in the lower back. However, like any surgical procedure, it carries certain risks and potential complications. It is important to discuss these risks with your surgeon before undergoing the surgery. Here are some potential risks of lumbar decompression surgery:

(1) Infection: There is a risk of developing a surgical site infection, which can occur at the incision site or within the deeper tissues. Infections may require additional treatment, such as antibiotics or drainage.
(2) Bleeding: While uncommon, there is a risk of excessive bleeding during or after surgery. In some cases, a blood transfusion might be necessary.
(3) Nerve injury: Surgery involves working near nerves, and there is a small risk of damaging adjacent nerves, which could result in new or worsened neurological symptoms.
(4) Dural tear: The dura is a thin, protective membrane surrounding the spinal cord and nerves. During surgery, the dura can accidentally be torn, resulting in cerebrospinal fluid leakage, which may require additional treatment.
(5) Persistent or worsened symptoms: In some cases, surgery may not completely relieve symptoms, or there could be temporary worsening of symptoms immediately after surgery due to inflammation or nerve irritation.
(6) Blood clots: Prolonged immobility during and after surgery can increase the risk of developing blood clots, which can be dangerous if they travel to the lungs (pulmonary embolism) or other parts of the body.
(7) Anesthesia complications: There are risks associated with anesthesia, including allergic reactions, breathing difficulties, and adverse reactions to medications.
(8) Instrumentation failure: If hardware (such as screws or rods) is used to stabilize the spine, there is a risk of instrumentation failure, including loosening, breakage, or migration.

Attending 1 Score: 3
Attending 2 Score: 3

Analysis: The response provided a somewhat useful, albeit cursory, overview of the potential complications involved in PLD. These complications are often addressed or raised in informed consent conversations with patients prior to surgery and, therefore, may serve as a useful introduction to these issues prior to an office visit. However, the mention of "instrumentation failure", among other complications, might confuse patients regarding the use of hardware in a simple, uninstrumented PLD. Additionally, the response does not provide any clarity regarding the actual prevalence of these adverse events for patients undergoing PLD, which would provide crucial context to anyone considering surgery.

5. What is the recovery time after lumbar decompression surgery?

Here's a general timeline for recovery after lumbar decompression surgery:

Immediate postoperative period (days to weeks):
- Hospital stay: Most patients stay in the hospital for a short period, usually a day or two, after lumbar decompression surgery.
- Pain management: Pain and discomfort are common after surgery. Pain medication and other methods are used to manage pain.
- Mobility: Patients are encouraged to start walking and moving around shortly after surgery to prevent complications like blood clots.

First two months:
- Gradual increase in activity: Patients can gradually increase their activity level under the guidance of their healthcare provider.
- Physical therapy: Physical therapy may be recommended to help improve strength, flexibility, and overall mobility.
- Return to work: Depending on the type of work, patients may be able to return to work within several weeks to a few months, but this varies.

Long-term:
- Full recovery: It can take up to a year or more to fully recover and achieve the maximum benefits of the surgery.
- Activity and lifestyle: Patients are often able to return to their normal activities and routines, though some modifications may be necessary to prevent future issues.

Attending 1 Score: 2
Attending 2 Score: 4

Analysis: Attendings differed in their opinion of the adequacy of Chat GPT's response to this question. It was acknowledged that this response was a good overview of a recovery timeline after PLD, while noting that it might mislead patients by not mentioning that many patients can be discharged on the day of surgery[30]. Another deficiency was the response's lack of specific information regarding patient activities postoperatively, which is one of the main concerns of patients considering surgery. Additionally, the response was seen as vague regarding return to work and potential restrictions. Lastly, the full recovery timeline was perceived to be inaccurate, as many patients return to full functionality well before a year after PLD[31].

6. How will lumbar decompression surgery affect my range of motion?
Lumbar decompression surgery can have varying effects on your range of motion depending on the specific procedure performed, the extent of the decompression, and individual factors. Here is a general overview of how lumbar decompression surgery might affect your range of motion:

(1) Minimally invasive approaches: Many minimally invasive lumbar decompression surgeries aim to minimize disruption to surrounding tissues and muscles. As a result, these procedures can potentially lead to less alteration in your range of motion compared to traditional open surgeries.
(2) Traditional open surgery: Traditional open lumbar decompression surgery might involve larger incisions and more tissue disruption. There might be some impact on the flexibility and range of motion of the spine in the immediate postoperative period.
(3) Microdiscectomy: If your surgery is focused on removing a herniated disc or portion of a disc, the impact on range of motion might be minimal.

(4) Laminectomy or laminotomy: These procedures involve removing part of the lamina (a bony arch in the spine) to relieve pressure on the spinal cord or nerves. While this can increase space for nerve structures, it might have a slight impact on stability and range of motion.

(5) Fusion procedures: If fusion is performed in conjunction with decompression, it might lead to a reduction in the range of motion at the fused segment.

It is important to note that your surgeon will strive to achieve a balance between decompression and preserving spinal stability and function. They will discuss the potential impact on your range of motion during the preoperative consultation, taking into consideration your specific condition, treatment goals, and lifestyle.

Attending 1 Score: 4
Attending 2 Score: 2

Analysis: Attendings disagreed on the adequacy of ChatGPT's response to this question. A lower score from one attending was given because ChatGPT's claim regarding the superiority of MISS in preserving range of motion is not conclusively supported by current spine surgery literature[32]. It was also felt that the mention of fusion could confuse patients considering PLD, which is typically handled without a concomitant fusion procedure. It was concluded that the response provided a satisfactory but uncomprehensive outline of motion preservation for different techniques and indications for PLD.

7. Will lumbar decompression surgery help my back pain?
Lumbar decompression surgery is primarily aimed at relieving symptoms caused by nerve compression, such as leg pain, weakness, numbness, and other neurological symptoms. While it can lead to a significant reduction or elimination of leg-related symptoms, its impact on back pain itself can vary. Here are some key points to consider:

(1) Secondary improvement: Lumbar decompression surgery is primarily designed to alleviate symptoms related to nerve compression. If your back pain is directly related to nerve compression or irritation, surgery might lead to a reduction in back pain as well.

(2) Mixed results: The impact of lumbar decompression surgery on isolated back pain (pain that is not primarily due to nerve compression) can be less predictable. Some patients experience relief, while others might have varying degrees of improvement.

(3) Degenerative disc disease: If your back pain is due to degenerative disc disease and is not primarily nerve-related, lumbar decompression surgery might not target the underlying cause of your pain. Other surgical options, like spinal fusion, might be considered in such cases.

(4) Patient expectations: It is important to have realistic expectations about the outcomes of surgery. Discuss your goals and expectations with your surgeon during your consultations.

Attending 1 Score: 4
Attending 2 Score: 2

Analysis: The reviewers generally agreed regarding the accuracy of the content provided in this response. While the response was seen as useful for patients considering PLD, it was noted that it also could be perplexing for patients without a medical background who might have numerous follow-up questions regarding the difference between "back pain" *vs.* "nerve pain". While many medical professionals can identify the difference between such etiologies, some patients may not be able to differentiate between these sources of pain. Therefore, patients could be confused as to how PLD might address their specific

symptoms. Another major deficiency in this response is the lack of data regarding outcomes after surgery, which contributed to a reduction in the overall score of the response.

8. What is the success rate of lumbar decompression surgery?

The success rate of lumbar decompression surgery can vary based on several factors, including the specific condition being treated, the underlying cause of symptoms, the surgical technique used, the patient's overall health, and the criteria used to define success. Here is a general overview:

(1) Herniated disc: Microdiscectomy or other decompression surgeries for herniated discs have generally high success rates, often above 80%-90%.
(2) Spinal stenosis: Success rates vary, but many studies report favorable outcomes in terms of symptom relief.
(3) Foraminal stenosis: Surgery for foraminal stenosis, which involves widening the nerve exit openings, can lead to relief of leg pain and other symptoms. Success rates vary but can be quite favorable.
(4) Degenerative disc disease: Success rates are often good, but outcomes can vary based on factors like the extent of degeneration and overall health.
(5) Spondylolisthesis: Success rates depend on factors like the grade of slippage and the success of fusion. Overall, many patients experience symptom improvement.
(6) Recurrent symptoms: While many patients experience relief after lumbar decompression surgery, there is a small risk of recurrent symptoms due to factors such as scar tissue formation, incomplete decompression, or new issues.

Attending 1 Score: 2
Attending 2 Score: 2

Analysis: This response was generally of poor quality and of little use to patients. The reviewers both felt that the response did not delineate the differing success rates of PLD for any of the indications listed after "herniated disk". Secondarily, the response repeats the same general answer for almost every indication - some of which, including spondylolisthesis, are typically treated by spinal fusion and not a standalone PLD. This could be seen as a point of frustration on the side of the patient as they begin to ask deeper questions from ChatGPT but are met with a plateauing effect of new information without specific resources to address lingering questions.

9. What are the different approaches for lumbar decompression surgery?

There are several surgical approaches for lumbar decompression, each designed to address specific conditions and anatomical considerations. The choice of approach depends on factors such as the underlying condition, the level of the spine affected, the surgeon's expertise, and patient characteristics. Here are some common approaches for lumbar decompression surgery:

(1) Traditional open approach:
- Involves a larger incision and more tissue dissection.
- Provides direct visualization of the surgical area.
- Allows for comprehensive decompression and treatment of complex conditions.
- Examples include laminectomy, laminotomy, and discectomy.

(2) Minimally invasive approach:
- Utilizes smaller incisions and specialized instruments.

- Causes less tissue disruption and reduces postoperative pain.
- Often leads to faster recovery and shorter hospital stays.
- Different techniques include tubular retractor, endoscopic, and microendoscopic approaches.

Attending 1 Score: 3
Attending 2 Score: 2

Analysis: While the complete response offered a broad list of potential approaches to PLD, the descriptions for many approaches were incomplete and misleading. For example, the reviewers felt that in the description of a corpectomy, ChatGPT neglected to mention that after the removal of a vertebral body, some patients may require cage placement to restore height. Additionally, several of the approaches, such as facetectomy and corpectomy, would not be used for a simple PLD and would necessitate concomitant fusion. The claims regarding some of the advantages of MISS were felt to not be fully representative of the literature[33-35]. As such, these responses were viewed as not wholly accurate and possibly confusing to patients considering their options before surgery.

10. Which approach is better for lumbar decompression?

The choice of approach for lumbar decompression surgery depends on various factors, including your specific condition, the underlying cause of your symptoms, the surgeon's expertise, and your individual needs. There are different surgical approaches that can be used for lumbar decompression, and what might be considered "better" depends on your unique situation.

- Condition: The type and severity of your spinal condition play a significant role in determining the appropriate approach.
- Health status: Your overall health, medical history, and any existing medical conditions might impact which approach is safer for you.
- Surgical expertise: The surgeon's experience and expertise with a particular approach are important factors.
- Patient preference: Discuss your goals, concerns, and expectations with your surgeon to choose an approach that aligns with your needs.
- Recovery goals: Consider your desired recovery timeline and how much downtime you can afford.

Attending 1 Score: 3
Attending 2 Score: 3

Analysis: This response provides a cursory overview of different approaches. However, it fails to provide evidence-based advantages or drawbacks for the different techniques. In addition, the responses are vague and do not sufficiently address the question or the nuances thereof. As such, it lacks both the specificity and evidence-based support to be helpful to patients considering surgery. Like the prior responses, there appears to be a plateauing of new information, even with different questions surrounding the same topic.

Average Attending Score: 2.95

DISCUSSION

The advent of online search engines has incited a change in how patients interact with medical information and, therefore, their physicians. New, open-access AI chatbot technology represents a revolutionary moment in medicine regarding how patients learn about their medical conditions. Given the potential

ramifications of such programs and tools, it is crucial to evaluate ChatGPT's utility and accuracy in disseminating orthopedic information. Chatbot responses may impact patients' perceptions of treatment options and risks prior to an evaluation by a physician. Several studies have analyzed the utility of ChatGPT for patients considering orthopedic surgery[11-17]. Assessing ChatGPT's usefulness for preoperative patient education in spine surgery is especially critical due to the relatively high risk of spine surgery and the nuances that often guide decision making regarding the indications for different operations. To our knowledge, the present study is the first to use a modified validated scoring system to appraise and evaluate ChatGPT's responses to common patient questions when considering PLD surgery.

Minimum scores across all ten questions would lead to a total score of 20, whereas a maximum score would be 100. ChatGPT's responses in this analysis earned a score of 59, just under an average score of 3, when evaluated by two attending, fellowship-trained orthopedic spine surgeons. A score of 3 denoted a somewhat useful response of moderate quality, with some important information adequately discussed but some poorly discussed [Figure 1].

In the present study, ChatGPT was generally able to provide an accurate, albeit cursory, overview of relevant surgical indications, techniques, complications, and alternate therapies. However, some of these answers, when evaluated individually, lacked the clarification necessary to provide patients with a thorough understanding to inform their medical decision making. Some of the answers have the potential to be harmful to patients, especially those answers suggesting alternative therapy without the necessary context of the patient's particular history and symptom severity. In some instances, for example, PLD might be necessary to reverse or prevent further neurologic injury, especially for urgent and emergent indications. Suggesting alternative, non-operative treatment options for these patients could worsen or adversely impact patient outcomes. Concordantly, a prior study reported that ChatGPT had a 53% mismanagement rate, which would be especially deleterious for serious underlying pathology[36]. Furthermore, non-operative treatment option descriptions were often vague, such as physical therapy to "strengthen muscles". This could lead some patients to pursue inadequate or harmful treatment, which may exacerbate or accelerate their disease processes.

Additionally, several of the claims were not fully substantiated by current spine surgery literature and several of the listed indications (spondylolisthesis and degenerative disc disease) may be better treated with other procedures, such as spinal fusion. As noted in previous literature, ChatGPT has been trained to generate definitive responses to questions, even when the existing literature may not be conclusive enough to make a specific recommendation[37,38]. In particular, the chatbot seemed to indicate the superiority of MISS over the traditional open approach. While there is increasing research regarding the potential benefits of minimally invasive surgery, there are still gaps in the literature, which can be most appropriately addressed by a trained and experienced surgeon[33,34]. These discrepancies may be confusing to patients considering PLD and could potentially lead to a delay in care. Nevertheless, ChatGPT did repeatedly emphasize that its responses should be taken in conjunction with consultation with a spine surgeon. This inability to address appropriate, patient-specific context affirms the findings of previous literature supporting the spine surgeon's role in providing individualized clinical recommendations[36].

One limitation of any study attempting to characterize the utility of online sources of medical information to patients prior to a doctor's visit is the inherent subjectivity with which the online source is evaluated. To combat this weakness, the present analysis implemented a more objective, validated numeric scoring system. Additionally, the responses were analyzed by two attending spine surgeons, both of whose scores were presented, providing additional insight from physicians with differing levels of experience and areas of

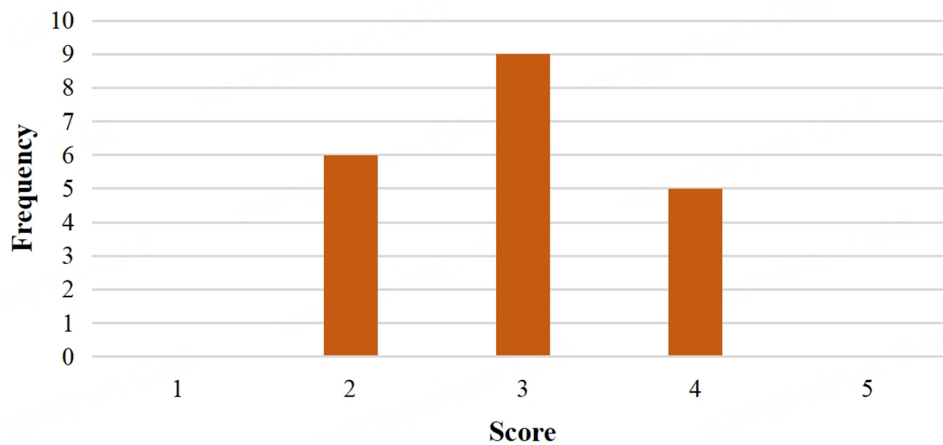## Summary of Attending Scores to ChatGPT Responses

Figure 1. Summary of scores to ChatGPT responses.

focus. A current limitation of ChatGPT software is its inability to provide patient-specific recommendations based on individual clinical data. One way to improve ChatGPT's software, therefore, would be to include a method by which a patient could upload imaging in a HIPPA-compliant manner. ChatGPT could then use this diagnostic information, in conjunction with patient-reported symptoms, to make more specific recommendations. Finally, the present study attempts to analyze ChatGPT's responses to common patient questions, as it is the most widely utilized and easily accessible AI chatbot presently available for consumers. However, future studies could aim to compare the clinical accuracy and utility of responses given by different chatbots to systematically determine which might be best suited to providing patients with useful preoperative information.

In summation, ChatGPT has the ability to provide broadly useful responses to common preoperative questions that patients may have when considering undergoing PLD. ChatGPT has excellent utility in providing background information to patients and in helping them become more informed about their pathology in general. However, it often lacks the specific patient context, especially patient imaging data such as X-ray, CT, and MRI, necessary to provide patients with proper, accurate insights into their personal prognosis and medical options. Fortunately, ChatGPT does state its own limitations and consistently recommends using any information it provides in context with the consultation of a spine surgeon. In such instances, where patients use ChatGPT as background information to prepare for in-person appointments with a spine surgeon, this additional knowledge can be beneficial in enabling them to actively participate in the shared decision-making process regarding their medical care and take shared responsibility for their outcomes.

DECLARATIONS
Authors' contributions
Made substantial contributions to the conception and design of the study and performed data analysis and interpretation: Giakas AM, Narayanan R, Ezeonu T, Dalton J, Mangan J, Schroeder G, Kepler C
Performed data acquisition, as well as providing administrative, technical, and material support: Giakas AM, Narayanan R, Ezeonu T, Mangan J, Schroeder G, Kepler C

Conceptualization, methodology, project administration, writing - original draft: Lee Y
Conceptualization, methodology, project administration, writing - review and editing: Dalton J
Formal analysis, investigation, visualization, writing - original draft: Henry T
Conceptualization, project administration, supervision, writing - review and editing: Vaccaro A

Availability of data and materials
Not applicable.

Financial support and sponsorship
None.

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Copyright
© The Author(s) 2024.

REFERENCES

1. Van Riel N, Auwerx K, Debbaut P, Van Hees S, Schoenmakers B. The effect of Dr Google on doctor-patient encounters in primary care: a quantitative, observational, cross-sectional study. *BJGP Open* 2017;1:bjgpopen17X100833. DOI PubMed PMC
2. Cocco AM, Zordan R, Taylor DM, et al. Dr Google in the ED: searching for online health information by adult emergency department patients. *Med J Aust* 2018;209:342-7. DOI PubMed
3. Fraval A, Ming Chong Y, Holcdorf D, Plunkett V, Tran P. Internet use by orthopaedic outpatients - current trends and practices. *Australas Med J* 2012;5:633-8. DOI PubMed PMC
4. Kasthuri V, Homer A, Alsoof D, et al. Modern internet search analytics and spine: what are patients asking and reading online? *N Am Spine Soc J* 2023;14:100214. DOI PubMed PMC
5. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell* 2023;6:1169595. DOI PubMed PMC
6. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare* 2023;11:887. DOI PubMed PMC
7. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *J Med Syst* 2023;47:33. DOI PubMed PMC
8. OpenAI. Introducing ChatGPT. 2022. Available from: https://openai.com/blog/chatgpt. [Last accessed on 27 Aug 2024].
9. Lock S. What is AI chatbot phenomenon ChatGPT and could it replace humans? 2022. Available from: https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans. [Last accessed on 27 Aug 2024].
10. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ* 2023;9:e45312. DOI PubMed PMC
11. Kaarre J, Feldt R, Keeling LE, et al. Exploring the potential of ChatGPT as a supplementary tool for providing orthopaedic information. *Knee Surg Sports Traumatol Arthrosc* 2023;31:5190-8. DOI PubMed PMC
12. Dubin JA, Bains SS, Chen Z, et al. Using a Google web search analysis to assess the utility of ChatGPT in total joint arthroplasty. *J Arthroplasty* 2023;38:1195-202. DOI PubMed
13. Duey AH, Nietsch KS, Zaidat B, et al. Thromboembolic prophylaxis in spine surgery: an analysis of ChatGPT recommendations. *Spine J* 2023;23:1684-91. DOI PubMed
14. Fayed AM, Mansur NSB, de Carvalho KA, Behrens A, D'Hooghe P, de Cesar Netto C. Artificial intelligence and ChatGPT in orthopaedics and sports medicine. *J Exp Orthop* 2023;10:74. DOI PubMed PMC
15. Hodakowski AJ, McCormick JR, Damodar D, et al. Rotator cuff repair: what questions are patients asking online and where are they getting their answers? *Clin Shoulder Elb* 2023;26:25-31. DOI PubMed PMC
16. Hurley ET, Crook BS, Lorentz SG, et al. Evaluation high-quality of information from ChatGPT (artificial intelligence-large language

model) artificial intelligence on shoulder stabilization surgery. *Arthroscopy* 2024;40:726-31.e6. DOI PubMed

17. Mika AP, Martin JR, Engstrom SM, Polkowski GG, Wilson JM. Assessing ChatGPT responses to common patient questions regarding total hip arthroplasty. *J Bone Joint Surg Am* 2023;105:1519-26. DOI PubMed

18. Subramanian T, Shahi P, Araghi K, et al. Using artificial intelligence to answer common patient-focused questions in minimally invasive spine surgery. *J Bone Joint Surg Am* 2023;105:1649-53. DOI PubMed

19. Lattig F, Fekete TF, O'Riordan D, et al. A comparison of patient and surgeon preoperative expectations of spinal surgery. *Spine* 2013;38:1040-8. DOI PubMed

20. Deyo RA, Mirza SK, Martin BI, Kreuter W, Goodman DC, Jarvik JG. Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA* 2010;303:1259-65. DOI PubMed PMC

21. O'Lynnger TM, Zuckerman SL, Morone PJ, Dewan MC, Vasquez-Castellanos RA, Cheng JS. Trends for spine surgery for the elderly: implications for access to healthcare in North America. *Neurosurgery* 2015;77 Suppl 4:S136-41. DOI PubMed

22. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. *Am J Gastroenterol* 2007;102:2070-7. DOI PubMed

23. Kreiner DS, Hwang SW, Easa JE, et al; North American Spine Society. An evidence-based clinical guideline for the diagnosis and treatment of lumbar disc herniation with radiculopathy. *Spine J* 2014;14:180-91. DOI PubMed

24. Zaina F, Tomkins-Lane C, Carragee E, Negrini S. Surgical versus non-surgical treatment for lumbar spinal stenosis. *Cochrane Database Syst Rev* 2016;2016:CD010264. DOI PubMed PMC

25. Katz JN, Zimmerman ZE, Mass H, Makhni MC. Diagnosis and management of lumbar spinal stenosis: a review. *JAMA* 2022;327:1688-99. DOI PubMed

26. Kuris EO, McDonald CL, Palumbo MA, Daniels AH. Evaluation and management of cauda equina syndrome. *Am J Med* 2021;134:1483-9. DOI PubMed

27. Bulloch L, Thompson K, Spector L. Cauda equina syndrome. *Orthop Clin North Am* 2022;53:247-54. DOI PubMed

28. Issack PS, Cunningham ME, Pumberger M, Hughes AP, Cammisa FP Jr. Degenerative lumbar spinal stenosis: evaluation and management. *J Am Acad Orthop Surg* 2012;20:527-35. DOI PubMed

29. Fritz JM, Delitto A, Welch WC, Erhard RE. Lumbar spinal stenosis: a review of current concepts in evaluation, management, and outcome measurements. *Arch Phys Med Rehabil* 1998;79:700-8. DOI PubMed

30. Goacher E, Sanders MI, Ivanov M. Safety and feasibility of same-day discharge following lumbar decompression surgery: a systematic review. *Brain Spine* 2022;2:100888. DOI PubMed PMC

31. Degen T, Fischer K, Theiler R, et al. Outcomes after spinal stenosis surgery by type of surgery in adults aged 60 years and older. *Swiss Med Wkly* 2020;150:w20325. DOI PubMed

32. Matsumoto K, Shah A, Kelkar A, et al. Biomechanical evaluation of a novel decompression surgery: transforaminal full-endoscopic lateral recess decompression (TE-LRD). *N Am Spine Soc J* 2021;5:100045. DOI PubMed PMC

33. Hathi K, Bigney E, Richardson E, et al. Minimally invasive vs. open surgery for lumbar spinal stenosis in patients with diabetes - a Canadian spine outcomes and research network study. *Global Spine J* 2023;13:1602-11. DOI PubMed PMC

34. Goldstein CL, Macwan K, Sundararajan K, Rampersaud YR. Perioperative outcomes and adverse events of minimally invasive versus open posterior lumbar fusion: meta-analysis and systematic review. *J Neurosurg Spine* 2016;24:416-27. DOI PubMed

35. Skovrlj B, Belton P, Zarzour H, Qureshi SA. Perioperative outcomes in minimally invasive lumbar spine surgery: a systematic review. *World J Orthop* 2015;6:996-1005. DOI PubMed PMC

36. Chalhoub R, Mouawad A, Aoun M, et al. Will ChatGPT be able to replace a spine surgeon in the clinical setting? *World Neurosurg* 2024;185:e648-52. DOI PubMed

37. Mejia MR, Arroyave JS, Saturno M, et al. Use of ChatGPT for determining clinical and surgical treatment of lumbar disc herniation with radiculopathy: a North American spine society guideline comparison. *Neurospine* 2024;21:149-58. DOI PubMed PMC

38. Rajjoub R, Arroyave JS, Zaidat B, et al. ChatGPT and its role in the decision-making for the diagnosis and treatment of lumbar spinal stenosis: a comparative analysis and narrative review. *Global Spine J* 2024;14:998-1017. DOI PubMed PMC