GPT-4 as a Source of Patient Information for Anterior Cervical Discectomy and Fusion: A Comparative Analysis Against Google Web Search

Paul G. Mastrokostas, BA, BBA¹, Leonidas E. Mastrokostas, BA², Ahmed K. Emara, MD³, Ian J. Wellington, MD⁴, Elizabeth Ginalis, MD⁵, John K. Houten, MD⁶, Amrit S. Khalsa, MD⁷, Ahmed Saleh, MD⁸, Afshin E. Razi, MD⁸, and Mitchell K. Ng, MD⁸

Abstract

Study Design: Comparative study.

Objectives: This study aims to compare Google and GPT-4 in terms of (1) question types, (2) response readability, (3) source quality, and (4) numerical response accuracy for the top 10 most frequently asked questions (FAQs) about anterior cervical discectomy and fusion (ACDF).

Methods: “Anterior cervical discectomy and fusion” was searched on Google and GPT-4 on December 18, 2023. Top 10 FAQs were classified according to the Rothwell system. Source quality was evaluated using JAMA benchmark criteria and readability was assessed using Flesch Reading Ease and Flesch-Kincaid grade level. Differences in JAMA scores, Flesch-Kincaid grade level, Flesch Reading Ease, and word count between platforms were analyzed using Student’s t-tests. Statistical significance was set at the .05 level.

Results: Frequently asked questions from Google were varied, while GPT-4 focused on technical details and indications/management. GPT-4 showed a higher Flesch-Kincaid grade level (12.96 vs 9.28, \( P = .003 \)), lower Flesch Reading Ease score (37.07 vs 54.85, \( P = .005 \)), and higher JAMA scores for source quality (3.333 vs 1.800, \( P = .016 \)). Numerically, 6 out of 10 responses varied between platforms, with GPT-4 providing broader recovery timelines for ACDF.

Conclusions: This study demonstrates GPT-4’s ability to elevate patient education by providing high-quality, diverse information tailored to those with advanced literacy levels. As AI technology evolves, refining these tools for accuracy and user-friendliness remains crucial, catering to patients’ varying literacy levels and information needs in spine surgery.

Keywords
GPT-4, Google, artificial intelligence, anterior cervical discectomy and fusion, anterior cervical discectomy and fusion, patient education, health literacy, readability

¹ College of Medicine, State University of New York (SUNY) Downstate, Brooklyn, NY, USA
² Brooklyn College of the City University of New York, Brooklyn, NY, USA
³ Department of Orthopaedic Surgery, Cleveland Clinic, Cleveland, OH, USA
⁴ Department of Orthopaedic Surgery, University of Connecticut, Hartford, CT, USA
⁵ Department of Neurosurgery, Rutgers University, Newark, NJ, USA
⁶ Department of Neurosurgery, Mount Sinai School of Medicine, New York, NY, USA
⁷ Department of Orthopaedic Surgery, University of Pennsylvania, Philadelphia, PA, USA
⁸ Department of Orthopaedic Surgery, Maimonides Medical Center, Brooklyn, NY, USA

Corresponding Author:
Paul G. Mastrokostas, BA, BBA, College of Medicine, State University of New York (SUNY) Downstate, 450 Clarkson Avenue, Brooklyn, New York 11203, USA.
Email: Pmastrokostas06@gmail.com

Creative Commons Non Commercial No Derivs CC BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License (https://creativecommons.org/licenses/by-nc-nd/4.0/) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
**Introduction**

ChatGPT, an (artificial intelligence) AI chatbot developed by OpenAI, is used by over 100 million weekly users for its ability to provide insightful answers to various queries. OpenAI released GPT-4 on March 13, 2023, enhancing ChatGPT’s capabilities; it is accessible through a ChatGPT Plus subscription. While both GPT-3.5, the current open-access AI chatbot of ChatGPT, and GPT-4 function on pattern recognition as a means of generating responses, GPT-4 has distinguished itself as a far more accurate and up-to-date source of information for a variety of user queries. GPT-4 actively pulls information from the Internet search engine Bing and reports sources when formulating responses to users’ questions, which is in stark contrast to the restricted bank of information that GPT-3.5 has access to. This technological advancement in AI chatbots, especially large language models (LLMs), brings new potential for their use in online patient education. Currently, 89% of U.S. citizens use Google for health-related information before consulting a physician. Google dominates the U.S. search engine market with 88.06% traffic, while Bing and Yahoo hold 6.94% and 2.70%, respectively. As such, Google was utilized as the metric for quality of questions and their corresponding answers were categorized using the Rothwell classification system, which divides questions into 3 categories: fact, policy, and value, further subclassified into 10 topics specific to ACDF (Table 1).

Linked websites corresponding to each question were evaluated for quality using the *Journal of the American Medical Association* (JAMA) benchmark criteria. These criteria, based on authorship, attribution, disclosure, and currency, have been validated in previous studies for assessing website quality. Two independent reviewers categorized these questions, and any disagreements were resolved by a senior author. The interobserver reliability for this classification was measured using Cohen’s kappa coefficients. Additionally, the top 10 FAQs with numerical responses from the Google search were inputted into GPT-4 for similar numerical outputs.

Comparative analysis of the responses from both platforms included assessments of readability, using the Flesch Reading Ease score and the Flesch-Kincaid grade level, and word count. The Flesch Reading Ease score measures readability on a scale from 0 to 100, utilizing a formula that incorporates average sentence length and average number of syllables per word. Higher scores indicate a higher degree of readability. Flesch Reading Ease scores of 90 to 100 typically indicate articles geared towards an elementary school student, scores of 60 to 70 are geared towards middle school students, and scores of 0 to 30 are geared towards university graduates. The Flesch Reading Ease score is thus inversely correlated with the Flesch-Kincaid Grade Level, with a higher score indicating a lower grade level. The Flesch Reading Ease score inversely correlates with the Flesch-Kincaid grade level. Plain-text format was used for analyzing shared responses (https://readability-score.com).

Cohen’s kappa coefficient was utilized to assess interobserver reliability using R statistical software (version 4.3.1; R Project for Statistical Computing, Vienna, Austria). Student’s t-tests were used to assess differences in the *JAMA* benchmark criteria scores, Flesch-Kincaid grade level, Flesch Reading Ease, and word count between GPT-4 and Google responses and website sources. Statistical significance was set at the .05 level.
Table 1. Description of Classification Systems for Questions and Websites.

<table>
<thead>
<tr>
<th>Rothwell's Classification</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fact</td>
<td>Asks whether something is true and to what extent.</td>
</tr>
<tr>
<td>Policy</td>
<td>Asks whether a specific course of action should be taken to solve a problem.</td>
</tr>
<tr>
<td>Value</td>
<td>Asks for an evaluation of an idea, object, or event.</td>
</tr>
</tbody>
</table>

Question classification by topic

Fact
Specific activities - Ability/ inability to perform a specific activity or action. | Example: Do you have to wear a neck brace after cervical fusion? |
Timeline of recovery - Length of time for recovery milestones. | Example: How long is recovery from anterior cervical discectomy and fusion? |
Technical details - Technical aspects of ACDF surgery, including instrumentation, approach, and anesthesia. | Example: How are the vertebrae fused together? |
Restrictions - Restrictions to activity or lifestyle in the recovery process. | Example: What are the permanent restrictions after cervical fusion? |
Cost - Cost of surgery/treatment. | Example: What is the cost of ACDF? |

Policy
Indications/Management - Specific indications for ACDF, as well as questions regarding the postoperative management of ACDF. | Example: Which conditions are treated with ACDF? |
Risks/Complications - Risks/complications occurring prior to, during, or following surgery, including the postoperative recovery and rehabilitation periods. | Example: What are the risks of ACDF? |

Value
Pain - Related to pain following ACDF. | Example: How much pain is there after anterior cervical discectomy and fusion? |
Longevity - Long-term complications and effects following ACDF. | Example: How long does ACDF last? |
Evaluation of surgery - Related to the successfulness, seriousness, or invasiveness of ACDF. | Example: Is ACDF considered major surgery? |

Website categorization
Commercial - Commercial organization that positions itself as a source of health information, including medical device companies. |
Academic - Academic institutions with a clear commitment to research and education, including universities, academic societies, and journals. |
Medical practice - Clinical practice or hospital with no clear academic affiliation. |
Government - Websites operated under the purview of a federal government, including URLs with gov and nhs.uk. |
Social media - Websites maintained by nonmedical organizations with the intent of spreading information between users, including blogs, internet forums, and videos uploaded to public entertainment platforms. |

Abbreviation: ACDF, Anterior cervical discectomy and fusion.

Results
Frequently Asked Questions in Google Web Search and GPT-4 for ACDF

Of the 10 most FAQs generated from a Google web search of ACDF, the most common Rothwell category was fact, accounting for 6 of the FAQs provided, followed by value questions, consisting of 4 questions (Table 2, Figure 1). There were no questions that were classified as policy. The distribution of Google’s top 10 most FAQs by topic was equal for specific activities, timeline of recovery, restrictions, pain, and evaluation of surgery, each accounting for 2 questions (Table 2, Figure 2). The most common Rothwell category for GPT-4 was fact, consisting of 5 questions. This was followed by policy, accounting for 4 questions (Table 2, Figure 1). One question was classified as value. The most common topics among questions produced by GPT-4 were technical details and indications/management, each accounting for 3 questions (Table 2, Figure 2). Interobserver reliability for Rothwell and topic categorizations were .918 and .942 ($P < .001$ for all), respectively, suggesting significant agreement between observers.
Google and GPT-4 shared one similar question regarding their top 10 FAQs, and the answers provided to this question varied (Table 3). Responses provided by Google had an average Flesch-Kincaid grade level of 9.28, an average Flesch Reading ease score of 54.85, and a mean word count of 158.30 (Table 4). Responses generated by GPT-4 had an average Flesch-Kincaid grade level of 12.96, Flesch Reading ease score of 37.07, and word count of 122.90. There was a statistically significant difference between Flesch-Kincaid grade level \((P = .003)\) and Flesch Reading ease score \((P = .005)\) between the answers provided via Google web search and GPT-4, with answers coming from GPT-4 showing a higher grade level and being more difficult to read. Overall, responses generated from GPT-4 were less direct and constituted a far longer timeline of recovery.

### Source Quality From Which Answers by Google Web Search and GPT-4 for FAQs Regarding ACDF Were Generated

Of the websites provided by Google web search to answer FAQs regarding ACDF, the most common source type was medical practice, accounting for 6 out of 10 unique websites, followed by social media and government, which comprised 3 out of 10 and 1 out of 10 unique websites, respectively. Of the unique websites utilized by GPT-4 to answer FAQs regarding ACDF, 2 out of 3 sources were categorized as social media, and one source was determined to be an academic website. Interobserver reliability for website categorization was .930 \((P < .001)\), suggesting high agreement between observers.

Source quality was further scrutinized through the assignment of JAMA scores. Of the unique websites that Google directed users to for answering FAQs, 6 out of 10 unique sources received a JAMA score of 1 (Table 4). 2 out of the 3 unique sources utilized by GPT-4 received a JAMA score of 3, and the remaining one source received a JAMA score of 4. The mean JAMA scores for sources utilized by Google Web search and GPT-4 to answer FAQs regarding ACDF were determined to be 1.800 and 3.333, respectively. A statistically significant difference \((P = .016)\) between mean JAMA scores of sources utilized by Google and GPT-4 was observed. Interobserver reliability for JAMA score categorization was 1.000 \((P < .001)\), suggesting complete agreement between observers.

### Numerical Responses in Google Web Search and GPT-4 for ACDF

The top 10 FAQs provided by a Google web search regarding ACDF with numerical responses were recorded and entered into GPT-4. Of the answers provided by Google and generated by GPT-4, 6 out of 10 questions varied in response (Table 5). Overall, GPT-4’s responses avoided specific numerical timeframes, favoring phrases such as “a few days,” “a few weeks,” and “several months” to convey durations.

### Table 2. Top 10 Most Frequently Asked Questions for ACDF.

<table>
<thead>
<tr>
<th>Google</th>
<th>GPT-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How serious is an anterior cervical discectomy and fusion? (Value/evaluation of surgery)</td>
<td>1. Is ACDF the right option for me? (Policy/indications or management)</td>
</tr>
<tr>
<td>2. How long is recovery from anterior cervical discectomy and fusion? (Fact/timeline of recovery)</td>
<td>2. What is a discectomy, and will I need this operation in addition to spinal fusion? (Fact/technical details)</td>
</tr>
<tr>
<td>3. How much pain is there after anterior cervical discectomy and fusion? (Value/pain)</td>
<td>3. How are the vertebrae fused together? (Fact/technical details)</td>
</tr>
<tr>
<td>4. Is ACDF considered major surgery? (Value/evaluation of surgery)</td>
<td>4. How long is the recovery from anterior cervical discectomy and fusion? (Fact/timeline of recovery)</td>
</tr>
<tr>
<td>5. How long do you sleep in a recliner after neck surgery? (Fact/timeline of recovery)</td>
<td>5. What are the risks of ACDF? (Policy/risks or complications)</td>
</tr>
<tr>
<td>6. What are the permanent restrictions after cervical fusion? (Fact/restrictions)</td>
<td>6. What is ACDF? (Fact/technical details)</td>
</tr>
<tr>
<td>7. What activities should you avoid after cervical fusion? (Fact/restrictions)</td>
<td>7. Which conditions are treated with ACDF? (Policy/indications or management)</td>
</tr>
<tr>
<td>8. How painful is cervical neck surgery? (Value/pain)</td>
<td>8. What can be expected on the day of surgery? (Value/evaluation of surgery)</td>
</tr>
<tr>
<td>9. Do you have to wear a neck brace after cervical fusion? (Fact/specific activities)</td>
<td>9. What is the recovery timeline following ACDF surgery? (Fact/timeline of recovery)</td>
</tr>
<tr>
<td>10. How do you sleep after cervical neck surgery? (Fact/specific activities)</td>
<td>10. Is ACDF right for me? (Policy/indications or management)</td>
</tr>
</tbody>
</table>

Abbreviation: ACDF, Anterior cervical discectomy and fusion.
Discussion

Large language models have been developed to serve a diverse array of functions, including assistance in education, recommendations for medical care, and individual case analysis in surgery.\(^{32,33}\) Despite having fewer parameters than other LLMs, ChatGPT is one of the most studied chatbots in the literature.\(^{34}\) The purpose of this study was to investigate GPT-4’s ability to provide patients with readable answers and reliable sources to FAQs related to ACDF in comparison to a Google web search conducted on December 18, 2023. While Google had an equal distribution of question topics, technical details and indications/management comprised the majority of FAQs queried through GPT-4. Answers provided by GPT-4 to FAQs were associated with a higher level of source evidence when compared to answers provided by a Google web search. Answers provided by GPT-4 were also associated with a higher Flesch-Kincaid grade level and lower Flesch Reading Ease score, yet similar word count, when compared to Google.

When asked to produce the 10 most FAQs for ACDF, GPT-4 only replicated 1 similar question that was queried through a Google web search. Currently, there is no available literature comparing the ability of GPT-4 to provide the 10 most FAQs related to a given search term to Google. Previous studies, like Dubin et al, found a 25% overlap in FAQs between GPT-3.5 and Google regarding hip and knee arthroplasty.\(^{17}\) GPT-4’s unique web search capabilities via Bing might contribute to these differences.\(^{8,9}\) Its method of determining FAQs could involve mining data from websites with dedicated FAQ sections, as opposed to querying Bing’s “People Also Ask” feature. Each GPT-4-generated question was linked to a

![Figure 1. Rothwell's classification stratified by search engine.](image)
website with a related question, enhancing the reliability of the information. The distinct approach of GPT-4 in generating FAQs, compared to traditional search engines, could influence how patients access and interpret information about ACDF, potentially altering their expectations and understanding of the procedure. This highlights the importance of evaluating the sources and methods of AI-generated medical information to ensure its alignment with evidence-based practices.

GPT-4 cited more reliable web sources than Google, which tended to rely on social media and medical practice websites. This is consistent with prior literature evaluating the most commonly directed websites for FAQs when querying Google for ACDF, whereas GPT-4 utilized sources such as academic and social media websites. The *JAMA* score of GPT-4’s sources was significantly higher, a novel finding given that earlier ChatGPT models didn’t provide source links, resulting in a *JAMA* score of zero. Walker et al also confirmed GPT-4’s ability to match the quality of existing online medical information, as assessed by the Ensuring Quality Information for Patients (EQIP) tool. Additionally, GPT-4’s reduced rate of citation fabrication marks a significant improvement when compared to GPT-3.5, reflecting a critical step forward in the utility of AI for providing trustworthy online health information. This shift towards more reliable and academically credible sources by GPT-4 suggests a notable enhancement in the quality of online health information provided by AI systems. It underscores the potential of AI tools like GPT-4 in supporting spine surgeons by directing patients to more evidence-based and peer-reviewed information, thereby improving patient education and potentially influencing treatment outcomes.

Answers provided by GPT-4 had a higher Flesch-Kincaid grade level and lower Flesch Reading Ease score than answers

![Figure 2. Topic classification stratified by search engine.](image-url)
provided by Google, aligning with earlier observations of GPT-3.5’s collegiate to post-collegiate readability levels.\textsuperscript{36,39,40} This suggests that GPT-4, like its predecessor, may not be fully accessible to the average American adult, whose health literacy is at or below an eighth-grade level (Flesch-Kincaid grade level 8) as defined by the National Institute of Health (NIH) and American Medical Association (AMA).\textsuperscript{41,42} However, ChatGPT can be prompted to simplify complex texts, significantly reducing collegiate reading level by an average of 3.3 grades.\textsuperscript{43} Despite GPT-4’s answers having a higher reading level, a previous study by Biro et al indicates that exposure to ChatGPT’s technical language responses can improve test scores, particularly among individuals with higher health literacy, and increase trust in the chatbot.\textsuperscript{44} The higher reading level of GPT-4’s responses, while potentially limiting accessibility for some, could also be seen as an opportunity to elevate patient understanding and engagement in their healthcare. This suggests the importance of tailoring AI-generated content to the reader’s literacy level, emphasizing the role of spine surgeons in interpreting and contextualizing information for patients to ensure comprehension and informed decision-making.

When comparing Google’s web search responses to those generated by GPT-4 for ACDF-related FAQs requiring numerical responses, 6 out of 10 questions had differing information between the 2 platforms. This finding aligns with a prior study on total joint arthroplasty, which reported that 55% of numerical question responses varied between the platforms.\textsuperscript{17} Google’s answers were more concise, while GPT-4 tended to offer a wider range of recovery timelines. For example, in response to “How long is recovery from ACDF?”, Google suggested 8-12 weeks, whereas GPT-4 provided an estimate ranging from a few weeks to 6 months. These

### Table 3. Answers to Similarly Asked Questions From Google and GPT-4 for ACDF.

<table>
<thead>
<tr>
<th>Question</th>
<th>Google</th>
<th>GPT-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google asked: How long is recovery from anterior cervical discectomy and fusion?</td>
<td>The usual period of recovery from cervical fusion surgery is 8 to 12 weeks and complete healing may take up to 6 months. Some patients may return to work sooner than others depending on their job, response to surgery, and ability to perform other lighter tasks in the workplace. Physician approval is required prior to returning to work.</td>
<td>The recovery time following anterior cervical discectomy and fusion (ACDF) surgery can vary but typically follows this general timeline: 1. Immediate post-Operative Period: Most patients spend 1 or 2 nights in the hospital. Some may even go home the same day. 2. Early recovery (First few weeks): Patients are advised to rest and limit their activities. A cervical collar or brace might be needed to support the neck. 3. Physical Therapy: Begins after a few weeks to restore strength, flexibility, and range of motion. This may continue for several weeks or months. 4. Return to Normal activities: Patients can often return to work and resume most everyday activities within 4 to 6 weeks. However, the time frame can vary based on the patient’s overall health and the nature of their work. 5. Complete fusion: For the fusion to fully mature and heal into 1 solid, strong bone, it may take 12 to 18 months. It’s important to follow your surgeon’s specific advice regarding activity restrictions and rehabilitation, as these will vary depending on individual health conditions and surgical outcomes.</td>
</tr>
<tr>
<td>GPT-4 asked: How long is the recovery from anterior cervical discectomy and fusion?</td>
<td>The recovery time following anterior cervical discectomy and fusion (ACDF) surgery can vary but typically follows this general timeline: 1. Immediate post-Operative Period: Most patients spend 1 or 2 nights in the hospital. Some may even go home the same day. 2. Early recovery (First few weeks): Patients are advised to rest and limit their activities. A cervical collar or brace might be needed to support the neck. 3. Physical Therapy: Begins after a few weeks to restore strength, flexibility, and range of motion. This may continue for several weeks or months. 4. Return to Normal activities: Patients can often return to work and resume most everyday activities within 4 to 6 weeks. However, the time frame can vary based on the patient’s overall health and the nature of their work. 5. Complete fusion: For the fusion to fully mature and heal into 1 solid, strong bone, it may take 12 to 18 months. It’s important to follow your surgeon’s specific advice regarding activity restrictions and rehabilitation, as these will vary depending on individual health conditions and surgical outcomes.</td>
<td>The recovery time following anterior cervical discectomy and fusion (ACDF) surgery can vary but typically follows this general timeline: 1. Immediate post-Operative Period: Most patients spend 1 or 2 nights in the hospital. Some may even go home the same day. 2. Early recovery (First few weeks): Patients are advised to rest and limit their activities. A cervical collar or brace might be needed to support the neck. 3. Physical Therapy: Begins after a few weeks to restore strength, flexibility, and range of motion. This may continue for several weeks or months. 4. Return to Normal activities: Patients can often return to work and resume most everyday activities within 4 to 6 weeks. However, the time frame can vary based on the patient’s overall health and the nature of their work. 5. Complete fusion: For the fusion to fully mature and heal into 1 solid, strong bone, it may take 12 to 18 months. It’s important to follow your surgeon’s specific advice regarding activity restrictions and rehabilitation, as these will vary depending on individual health conditions and surgical outcomes.</td>
</tr>
</tbody>
</table>

### Table 4. Comparison of Readability and Source Quality.

<table>
<thead>
<tr>
<th>Measures of Readability</th>
<th>Google X (SD)</th>
<th>GPT-4 X (SD)</th>
<th>P-Valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flesch-Kincaid grade level</td>
<td>9.28 (2.44)</td>
<td>12.96 (2.25)</td>
<td>P = .003</td>
</tr>
<tr>
<td>Flesch reading ease score</td>
<td>54.85 (14.15)</td>
<td>37.07 (10.25)</td>
<td>P = .005</td>
</tr>
<tr>
<td>Word count</td>
<td>158.30 (134.90)</td>
<td>122.90 (22.29)</td>
<td>P = .433</td>
</tr>
<tr>
<td>JAMA Score</td>
<td>N (%)</td>
<td>N (%)</td>
<td>P-Valuea</td>
</tr>
<tr>
<td>1</td>
<td>6 (60.0)</td>
<td>0 (0)</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>1 (10.0)</td>
<td>0 (0)</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>2 (20.0)</td>
<td>2 (66.7)</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>1 (10.0)</td>
<td>1 (33.3)</td>
<td>—</td>
</tr>
<tr>
<td>Mean JAMA score</td>
<td>1.800 (1.135)b</td>
<td>3.333 (.577)b</td>
<td>P = .016</td>
</tr>
</tbody>
</table>

X, mean; SD, standard deviation. 
aBolded P-value indicates statistical significance. 
bValue expressed as mean (standard deviation). 

ACDF, Anterior cervical discectomy and fusion.
variations in response timelines are crucial in the medical context, as they can influence patient expectations and preparedness. The broader spectrum of timelines from GPT-4 might reflect its capability to incorporate a wider array of patient experiences and clinical outcomes, as opposed to the more standardized responses typically found in Google searches. However, more objective questions, like “How long is hospital stay after cervical fusion?”, received consistent responses of 1-2 days from both platforms, indicating a consensus on certain clinical aspects. Overall, these findings suggest that while GPT-4 offers more comprehensive and varied responses, potentially enhancing patient understanding and expectations, it is also important to recognize the value of Google’s concise answers for certain straightforward, objective medical questions. This balance between detailed and straightforward information is essential in providing patients with a well-rounded understanding of their expectations following ACDF.

This study has several limitations. Google’s web search results and GPT-4’s responses, which are dependent on Bing, may vary due to updated content, affecting response accuracy. As a result, our findings might not be reproducible. Despite this limitation, this study provides a starting point for evaluating the capabilities of GPT-4 in answering FAQs relating to ACDF. Additionally, despite using a newly installed Google Chrome browser to avoid personalized results, Google might still have personalized the search outcomes. For GPT-4, user login requirements could introduce biases, as its responses might be influenced by the user’s interaction history. GPT-4 being a paid subscription service, unlike the open-access GPT-3.5, may bias its responses due its limited user base. The search methodology used to query GPT-4 in this study might not fully reflect the natural language of patients. Despite these challenges, this study provides important insights into GPT-4’s capabilities in medical information provision, highlighting the evolving role of AI in healthcare. Future studies should explore methods to prompt GPT-4 to respond in ways that more closely represent patient viewpoints.

### Conclusion

AI’s role in patient education is expanding, with GPT-4 showcasing enhanced capabilities and accuracy, though its higher reading level might limit usability for some users. This study demonstrates GPT-4’s potential to complement sources like Google, providing varied, in-depth responses. Yet, it emphasizes the need to balance complex AI content with clear, concise information. As AI integrates further into healthcare, refining these tools for accuracy and accessibility remains essential, meeting the diverse information needs and literacy levels of online medical information seekers.

### Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Mitchell K. Ng is a paid consultant at Ferghana Partners. For the remaining authors none were declared.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

<table>
<thead>
<tr>
<th>Question</th>
<th>Google</th>
<th>GPT-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How long is recovery from anterior cervical discectomy and fusion?</td>
<td>8-12 weeks</td>
<td>Few weeks, 1-6 months</td>
</tr>
<tr>
<td>2. How long do you sleep in a recliner after neck surgery?</td>
<td>1-2 days</td>
<td>Few days to several weeks</td>
</tr>
<tr>
<td>3. How long do you need to wear collar after cervical neck surgery?</td>
<td>6-8 weeks</td>
<td>4-6 weeks</td>
</tr>
<tr>
<td>4. How long do you need pain meds after cervical fusion?</td>
<td>1-4 weeks</td>
<td>Few days to several weeks</td>
</tr>
<tr>
<td>5. How long is hospital stay after cervical fusion?</td>
<td>1-2 days</td>
<td>1-2 days</td>
</tr>
<tr>
<td>6. How long does it take for neck fusion to fuse?</td>
<td>3 months</td>
<td>3-6 months</td>
</tr>
<tr>
<td>7. How long do you lose your voice after neck surgery?</td>
<td>Most people fully recover their voice within a few days of the surgery, but others may have trouble speaking for weeks or longer</td>
<td>Few days to a few weeks</td>
</tr>
<tr>
<td>8. How long will my throat hurt after neck surgery?</td>
<td>2-3 days</td>
<td>Few days to a week</td>
</tr>
<tr>
<td>9. How long does it hurt to swallow after ACDF!</td>
<td>At least 12 weeks</td>
<td>Several weeks to a few months</td>
</tr>
<tr>
<td>10. How long does it take to feel normal after ACDF surgery?</td>
<td>3 months</td>
<td>1 year or more</td>
</tr>
</tbody>
</table>

ACDF, Anterior cervical discectomy and fusion.

*a*Questions highlighted in bold varied in response.
References


