Observational Study

Medicine

OPEN

The assessment of the validity, safety, and utility of ChatGPT for patients with herniated lumbar disc

A preliminary study

Seoyon Yang, MD, PhD[a], Min Cheol Chang, MD[b,*]

Downloaded from http://journals.lww.com/md-journal by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 11/18/2024

Abstract

ChatGPT is perceived as a potential tool for patients diagnosed with herniated lumbar disc (HLD) to ask questions concerning desired information, with provision for necessary responses. In this preliminary study, we assessed the validity, safety, and utility of ChatGPT in patients with HLD. Two physicians specializing in the treatment of musculoskeletal disorders discussed and determined the 12 most frequently asked questions by patients with HLD in clinical practice. We used ChatGPT (version 4.0) to ask questions related to HLD. Each question was inputted into ChatGPT, and the responses were assessed by the 2 physicians. A Likert score was used to evaluate the validity, safety, and utility of the responses generated by ChatGPT. Each score for validity, safety, and utility was divided into 4 points, with a score of 4 indicating the most valid, safe, and useful answers and 1 point indicating the worst answers. Regarding validity, ChatGPT responses demonstrated 4 points for 9 questions (9/12, 75.0%) and 3 points for 3 questions (3/12, 25.0%). Regarding safety, ChatGPT scored 4 points for 11 questions (11/12, 91.7%) and 3 points for 1 question (1/12, 8.3%). Regarding utility, ChatGPT responses exhibited 4 points for 9 questions (9/12, 75.0%) and 3 points for 3 questions (3/12, 25.0%). ChatGPT demonstrates a tendency to offer relatively valid, safe, and useful information regarding HLD. However, users should exercise caution as ChatGPT may occasionally provide incomplete answers to some questions on HLD.

Abbreviation: HLD = herniated lumbar disc.

Keywords: artificial intelligence, ChatGPT, herniated lumbar disc, information, safety, utility, validity

1. Introduction

Herniated lumbar disc (HLD) is the displacement of the nucleus pulposus of the intervertebral disc due to a tear in the annulus fibrosus, extending beyond the inter-vertebral disc space into the spinal canal.[1] HLD is one of the most prevalent spinal disorders and can cause lower back and leg pains.[2] In the past, when patients experienced symptoms related to HLD, they were compelled to visit a hospital or a clinic to receive explanations from physicians regarding disease characteristics, treatment options, and prognosis. However, owing to limited accessibility to physicians and restricted clinical hours, many patients with HLD have difficulty obtaining sufficient medical information.

With the advent of the internet, patients can now obtain medical information on specific disorders or conditions online.[3,4] However, the Internet provides an overwhelming amount of information that extends beyond patients' specific queries. Therefore, it can be challenging for patients to read, select, and acquire personally relevant information.

Recently, a large language model, which is a sophisticated artificial intelligence model that excels in natural language processing tasks, has been developed.[5] It is trained using deep-learning techniques on massive amounts of Internet text data, allowing it to understand and respond to a wide range of topics.[6] ChatGPT is the most popular large language model and was developed by OpenAI based on the generative pretrained transformer architecture.[7–9] The primary function of ChatGPT is to provide human-like answers to natural language questions in real time.[7–9] This method is expected to be useful for applications in the medical field. We suppose that ChatGPT could serve as a tool for patients with HLD to ask questions related to the information they want to know, and that information can be provided by ChatGPT.[7–9] Thus, ChatGPT is expected to be a useful search engine for patients with HLD. However, the usefulness and accuracy of the information provided by ChatGPT

This work was supported by the National Research Foundation of Korea grant funded by the Korean government (MSIT) (NO.RS-2023-00219725).

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Supplemental Digital Content is available for this article.

[a] Department of Rehabilitation Medicine, School of Medicine, Ewha Woman's University Seoul Hospital, Seoul, Republic of Korea, [b] Department of Rehabilitation Medicine, College of Medicine, Yeungnam University, Daegu, Republic of Korea.

** Correspondence: Min Cheol Chang, Medicine, Yeungnam University 317-1, Daemyungdong, Namku, Daegu 705-717, Republic of Korea (e-mail: wheel633@hanmail.net).*

Copyright © 2024 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Yang S, Chang MC. The assessment of the validity, safety, and utility of ChatGPT for patients with herniated lumbar disc: A preliminary study. Medicine 2024;103:23(e38445).

Received: 25 January 2024 / Received in final form: 1 May 2024 / Accepted: 10 May 2024

http://dx.doi.org/10.1097/MD.0000000000038445

Table 1

Likert scores of each answer generated by ChatGPT.

Questions	Validity	Safety	Utility
1. What is a herniated lumbar disc?	4	4	4
2. What are the reasons for herniated lumbar discs?	4	4	4
3. What is the treatment for herniated lumbar discs?	3	4	3
4. When should patients with herniated lumbar discs undergo surgery?	4	4	4
5. Among nonsurgical and surgical treatments, which is better for patients with herniated lumbar discs?	4	4	4
6. Which postures or activities should patients with herniated lumbar discs avoid?	4	4	4
7. Which exercises are recommended for patients with herniated lumbar discs?	4	4	4
8. How many epidural steroid injections should be administered to patients with herniated lumbar disks?	3	3	3
9. Can herniated lumbar discs lead to disabilities?	4	4	4
10. What is the recurrence rate of herniated lumbar disks?	3	4	3
11. Is it advisable for a patient with an acute herniated lumbar disc to engage in daily activities or is it better to rest and lie down?	4	4	4
12. Is wearing a brace beneficial in patients with herniated lumbar disc?	4	4	4

Downloaded from http://journals.lww.com/md-journal by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 11/18/2024

should be evaluated with regards to HLD, as already done for other medical branches.[10–12]

Therefore, in this study, we assessed the validity, safety, and utility of the ChatGPT in providing information to patients with HLD seeking information about their condition.

2. Methods

2.1. Questions for asking to ChatGPT

Two authors (MCC and SY), each with approximately 15 years of experience in treating musculoskeletal disorders, discussed and determined the 12 most frequently asked questions by patients with HLD in clinical practice: (1) What is a herniated lumbar disc? (2) What are the reasons for herniated lumbar discs? (3) What is the treatment for herniated lumbar discs? (4) When should patients with herniated lumbar discs undergo surgery? (5) Among nonsurgical and surgical treatments, which is better for patients with herniated lumbar discs? (6) Which postures or activities should patients with herniated lumbar discs avoid? (7) Which exercises are recommended for patients with herniated lumbar discs? (8) How many epidural steroid injections should be administered to patients with herniated lumbar disks? (9) Can herniated lumbar discs lead to disabilities? (10) What is the recurrence rate of herniated lumbar disks? (11) Is it advisable for a patient with an acute herniated lumbar disc to engage in daily activities or is it better to rest and lie down? (12) Is wearing a brace beneficial in patients with herniated lumbar discs?

We used ChatGPT (version 4.0) to ask questions related to HLD. Each question was input into ChatGPT, and the responses were assessed by 2 authors (MCC and SY). Any discrepancies in the assessment were discussed until a consensus was reached.

Ethical approval was waived by the Institutional Review Board of Yeungnam University Hospital because human participants or animals were not involved for this study.

2.2. Assessment of the validity, safety, and utility of ChatGPT

A Likert score was used to evaluate the validity, safety, and utility of the responses generated by ChatGPT. The Likert scales used to evaluate validity, safety, and utility were categorized as follows:

- Validity:
  1. Completely erroneous information (all the information that ChatGPT answered cannot be found in medical sources or is inaccurate or incomplete).
  2. Partially erroneous information (some of the information that ChatGPT answered cannot be found in medical sources or contains inaccuracies or is incomplete).
  3. Reliable but incomplete information (all the information that ChatGPT answered is found in medical sources and is accurate but has some incomplete elements).
  4. Completely reliable and complete information (all the information that ChatGPT answered is found in medical sources and is complete).
- Safety:
  1. Significant and imminent danger to the patient's condition.
  2. Moderate potential danger to the patient's condition.
  3. Minimal potential danger to the patient's condition.
  4. No danger.
- Utility:
  1. Not useful for the patient (no useful information).
  2. Partially useful for the patient (>0% and <50% of the information provided is useful).
  3. Moderately useful for the patient (≥50% of the information provided is useful, but not 100%).
  4. Completely useful (100% of the information provided is useful).

3. Results

The Likert-scale scores for each question rated by the 2 experts (MCC and SYY) are presented in Table 1. The details of the answers provided by ChatGPT are provided in Data S1, Supplemental Digital Content, http://links.lww.com/MD/M735. Regarding validity, ChatGPT responses received 4 points for 9 questions (9/12, 75.0%; questions 1, 2, 4, 5, 6, 7, 9, 11, and 12) and 3 points for 3 questions (3/12, 25.0%; questions 3, 8, and 10). Regarding safety, ChatGPT responses scored 4 points for 11 questions (11/12, 91.7%; questions 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, and 12) and 3 points for 1 question (1/12, 8.3%; question 8). Regarding utility, ChatGPT responses received 4 points for 9 questions (9/12, 75.0%; questions 1, 2, 4, 5, 6, 7, 9, 11, and 12), and 3 points for 3 questions (3/12, 25.0%; questions 3, 8, and 10). Overall, ChatGPT's responses to questions regarding HLD appear valid, safe, and useful.

The validity and utility scores for Question 3 (What is the treatment for HLD?) were rated at 3 points because various treatment options, such as electrical therapy using transcutaneous electrical stimulation[13] and manual therapy,[14] have not been reported. Furthermore, ChatGPT reported a list of medications for HLD but did not include opioids, serotonin and norepinephrine reuptake inhibitors, and anticonvulsants, such as topiratmate, pregabalin, or gabapentin.[15–17] In addition, we believe that percutaneous epidural neuroplasty,[18] nucleoplasty,[19] or pulsed radiofrequency neuromodulation[20] could be considered as alternative procedures for treating HLD when steroid injections are ineffective.

Downloaded from http://journals.lww.com/md-journal by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 11/18/2024

The validity, safety, and utility of Question 8 (How many times should epidural steroid injections be administered to patients with HLD?) were rated at 3 points. Epidural steroid injections are recommended to be administered 3 to 4 times per year, with a minimum duration of at least 2 weeks between injections.[21] Multiple injections are often required to effectively reducing pain.[21] In addition, the possible side effects of epidural steroid injections, including nerve root damage, hematoma, transient non-positional headache, nausea, facial flushing, vasovagal reaction, and increased blood sugar in insulin-dependent diabetic patients, have not been fully described.[22]

The validity and utility of Question 10 (What is the recurrence rate of HLD?) were rated at 3 points. The possible factors that can induce the recurrence of HLD have been described in detail by ChatGPT. However, even though it is known to be approximately 15%,[23,24] the recurrence rate of HLD was not clearly answered.

4. Discussion

In this study, we assessed the validity, safety, and utility of ChatGPT for patients seeking information on HLD. Out of the 12 most commonly asked questions by patients with HLD in clinical practice, ChatGPT answered 75% of the questions with completely reliable and complete information, and 91.7% answers were found to be no danger to patients' condition. In addition, 75% of the answers were completely useful. Despite not achieving a perfect score, these answers were evaluated as having relatively satisfactory levels of validity, safety, and utility, with each receiving a score of 3. Therefore, we believe that ChatGPT answers questions regarding HLD as relatively valid, safe, and useful.

Despite favorable assessment results for validity, safety, and utility, there are some shortcomings in ChatGPT answers on HLD when compared to that of physicians. ChatGPT did not provide all the information about therapeutic options for HLD and failed to provide accurate details regarding the appropriate number of epidural steroid injections and the likelihood of HLD recurrence. However, it is realistically challenging for physicians to provide information on certain questions to patients. In addition, there is still no clear consensus on the appropriate frequency of epidural steroid injections, and the recurrence rates reported in previous studies are somewhat different according to the surveyed patient population.[25–27]

ChatGPT can be used as a search engine and source for obtaining medical information; however, it has some limitations, such as generating inaccurate or even erroneous responses.[28] Although ChatGPT may provide answers that are consistent and reliable, it can also provide false and misleading answers. For example, ChatGPT can be used in patients with common symptoms, such as fever,[29] where it can provide suggestions for taking antipyretics to relieve fever but it cannot accurately judge infection, pimples, or other causes. This can lead to an inaccurate diagnosis or delayed treatment.[29] However, among the answers provided by ChatGPT on HLD, despite some incomplete responses, none were found to be inaccurate nor erroneous.

ChatGPT has several issues.[30] First, ChatGPT cannot provide up-to-date medical information, because it is trained using data up to the year 2021. Second, the data that ChatGPT was trained to use may have been biased, because it could not distinguish between publications and articles. Third, ChatGPT may provide different answers to questions written at different times.

5. Conclusion

In conclusion, ChatGPT appears to provide relatively valid, safe, and useful information regarding HLD. ChatGPT can be used by patients to ask questions related to information that they want to know about HLD and obtain the necessary information. However, users should be aware that ChatGPT may provide incomplete responses to some questions on HLD. Our study has several limitations. We did not assess satisfaction with ChatGPT responses in patients with HLD. Additionally, the evaluations of validity, safety, and utility were relatively subjective. Further studies are warranted to address these limitations.

Author contributions

Conceptualization: Seoyon Yang, Min Cheol Chang.
Data curation: Seoyon Yang, Min Cheol Chang.
Formal analysis: Seoyon Yang, Min Cheol Chang.
Investigation: Seoyon Yang, Min Cheol Chang.
Methodology: Seoyon Yang, Min Cheol Chang.
Resources: Seoyon Yang, Min Cheol Chang.
Software: Min Cheol Chang.
Supervision: Min Cheol Chang.
Validation: Seoyon Yang, Min Cheol Chang.
Writing – original draft: Seoyon Yang, Min Cheol Chang.
Writing – review & editing: Seoyon Yang, Min Cheol Chang.

References

[1] Kwak SG, Choo YJ, Kwak S, Chang MC. Effectiveness of transforaminal, interlaminar, and caudal epidural injections in lumbosacral disc herniation: a systematic review and network meta-analysis. Pain Physician. 2023;26:113–23.
[2] Kwak S, Jang SH, Chang MC. Long-term outcomes of transforaminal epidural steroid injection in patients with lumbosacral radicular pain according to the location, type, and size of herniated lumbar disc. Pain Pract. 2021;21:836–42.
[3] Chang MC, Park D. YouTube as a source of information on epidural steroid injection. J Pain Res. 2021;14:1353–7.
[4] Lee H, Chang MC. YouTube as a source of information regarding the effect of vitamin C on coronavirus disease. Complement Ther Med. 2022;67:102827.
[5] Miao H, Ahn H. Impact of ChatGPT on interdisciplinary nursing education and research. Asian Pac Isl Nurs J. 2023;7:e48136.
[6] Clusmann J, Kolbinger FR, Muti HS, et al. The future landscape of large language models in medicine. Commun Med (Lond). 2023;3:141.
[7] Nedbal C, Naik N, Castellani D, Gauhar V, Geraghty R, Somani BK. ChatGPT in urology practice: revolutionizing efficiency and patient care with generative artificial intelligence. Curr Opin Urol. 2023;34:98–104.
[8] Ramamurthi A, Are C, Kothari AN. From ChatGPT to treatment: the future of AI and large language models in surgical oncology. Indian J Surg Oncol. 2023;14:537–9.
[9] Yu P, Xu H, Hu X, Deng C. Leveraging generative AI and large language models: a comprehensive roadmap for healthcare integration. Healthcare (Basel). 2023;11:2776.
[10] Frosolini A, Franz L, Benedetti S, et al. Assessing the accuracy of ChatGPT references in head and neck and ENT disciplines. Eur Arch Otorhinolaryngol. 2023;280:5129–33.
[11] Hatia A, Doldo T, Parrini S, et al. Accuracy and completeness of ChatGPT-generated information on interceptive orthodontics: a multicenter collaborative study. J Clin Med. 2024;13:735.
[12] Shrestha N, Shen Z, Zaidat B, et al. Performance of ChatGPT on NASS clinical guidelines for the diagnosis and treatment of low back pain: a comparison study. Spine (Phila Pa 1976). 2024;49:640–51.
[13] Sherry E, Kitchener P, Smart R. A prospective randomized controlled study of VAX-D and TENS for the treatment of chronic low back pain. Neurol Res. 2001;23:780–4.
[14] López-Díaz JV, Arias-Buría JL, Lopez-Gordo E, Lopez Gordo S, Oyarzún APA. "Effectiveness of continuous vertebral resonant oscillation using the POLD method in the treatment of lumbar disc hernia." A randomized controlled pilot study. Manual Therapy. 2015;20:481–6.
[15] Lee JH, Choi KH, Kang S, et al. Nonsurgical treatments for patients with radicular pain from lumbosacral disc herniation. Spine J. 2019;19:1478–89.
[16] Atkinson JH, Slater MA, Capparelli EV, et al. A randomized controlled trial of gabapentin for chronic low back pain with and without a radiating component. Pain. 2016;157:1499–507.
[17] Khoromi S, Patsalides A, Parada S, Salehi V, Meegan JM, Max MB. Topiramate in chronic lumbar radicular pain. J Pain. 2005;6:829–36.

[18] Taheri A, Khajenasiri AR, Nazemian Yazdi NA, Safari S, Sadeghi J, Hatami M. Clinical evaluation of percutaneous caudal epidural adhesiolysis with the racz technique for low back pain due to contained disc herniation. Anesth Pain Med. 2016;6:e26749.
[19] Yin H, Zhang X, Huang Z, Song Y, Zhu Q. Efficacy of single level versus double levels surgery of Percutaneous Disc Nucleoplasty (PDN) approach in treating lumbar disc herniation. Med Sci Monit. 2021;27:e930000.
[20] Shanthanna H, Chan P, McChesney J, Thabane L, Paul J. Pulsed radiofrequency treatment of the lumbar dorsal root ganglion in patients with chronic lumbar radicular pain: a randomized, placebo-controlled pilot study. J Pain Res. 2014;7:47–55.
[21] Hassan KZ, Sherman Al. Epidural Steroids. [Updated 2023 Jul 17]. In: StatPearls. Treasure Island (FL): StatPearls Publishing. 2023. Available at: https://www.ncbi.nlm.nih.gov/books/NBK537320/.
[22] Botwin KP, Gruber RD, Bouchlas CG, Torres-Ramos FM, Freeman TL, Slaten WK. Complications of fluoroscopically guided transforaminal lumbar epidural injections. Arch Phys Med Rehabil. 2000;81:1045–50.
[23] Beack JY, Chun HJ, Bak KH, Choi K-S, Bae I-S, Kim KD. Risk factors of secondary lumbar discectomy of a herniated lumbar disc after lumbar discectomy. J Korean Neurosurg Soc. 2019;62:586–93.
[24] Oh JT, Park KS, Jung SS, et al. Surgical results and risk factors for recurrence of lumbar disc herniation. Korean J Spine. 2012;9:170–5.
[25] Gaston P, Marshall RW. Survival analysis is a better estimate of recurrent disc herniation. J Bone Joint Surg Br. 2003;85:535–7.
[26] Jung JM, Lee SU, Hyun SJ, et al. Trends in incidence and treatment of herniated lumbar disc in Republic of Korea: a nationwide database study. J Korean Neurosurg Soc. 2020;63:108–18.
[27] Suk KS, Lee HM, Moon SH, Kim NH. Recurrent lumbar disc herniation: results of operative management. Spine (Phila Pa 1976). 2001;26:672–6.
[28] 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.
[29] Xue VW, Lei P, Cho WC. The potential impact of ChatGPT in clinical and translational medicine. Clin Transl Med. 2023;13:e1216.
[30] Uz C, Umay E. "Dr ChatGPT": is it a reliable and useful source for common rheumatic diseases? Int J Rheum Dis. 2023;26:1343–9.

Downloaded from http://journals.lww.com/md-journal by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC1y0abggQZXdtwnfKZBYtws= on 11/18/2024