

Research Articles

How Easy Is It to Get a Grip on Hand Pathologies? A Comparative Analysis of ChatGPT Against AAOS, AAHS, and ASSH Patient Information Sheets

Teren Yedikian, MD¹^a, Amanda Azer, MD¹, Hershil Patel, BS¹, Ariana Shaari, BS¹, Brian Molokwu, MS¹, Kush Mody, MD¹, Irfan Ahmed, MD¹, Michael Vosbikian, MD¹

¹ Department of Orthopaedics, Rutgers New Jersey Medical School, Newark, NJ, USA

Keywords: AAHS, AAOS, ASSH, ChatGPT, Upper Extremity

<https://doi.org/10.58616/001c.143190>

SurgiColl

Vol. 3, Issue 4, 2025

Objectives

Chat generative pre-trained transformer (ChatGPT) is a conversational artificial intelligence tool that can compose, analyze, and present information to its users. This comparative study aimed to explore whether ChatGPT can generate patient information sheets on common hand pathologies for the average US patient. The comprehensibility (dubbed “readability”) of ChatGPT was examined and compared to that of the American Academy of Orthopaedic Surgeons (AAOS), the American Association of Hand Surgery (AAHS), and the American Society for Surgery of the Hand (ASSH) patient information sheets.

Methods

Patient information sheets related to common hand pathologies were identified through the AAOS, AAHS, and ASSH websites. All entries used were pathologies limited to the hand and wrist. ChatGPT was utilized to generate patient information sheets on the same hand pathologies at the sixth-grade reading level. WebFx was utilized to calculate readability scores for the AAOS, AAHS, ASSH, and ChatGPT-generated patient sheets. Statistical analysis was determined using a paired two-tailed t-test. Statistical significance was defined as $P < 0.05$.

Results

Based on the results, patient information sheets from AAOS and ASSH are significantly easier to read on most metrics. Those from AAHS are no different from those generated from ChatGPT except for being easier to read per the Flesch-Kincaid Reading Ease and Coleman-Liau Index. Information sheets generated by ChatGPT have a significantly smaller number of words and complex words compared to those from AAOS and ASSH.

Conclusions

Patients may find it easier to read patient information sheets from AAOS and ASSH compared to those generated by ChatGPT.

INTRODUCTION

Artificial intelligence (AI) language models such as Chat generative pre-trained transformer (ChatGPT) have expe-

rienced a recent rise in popularity. ChatGPT, launched in fall 2022, is a conversational tool that can compose, analyze, and present information to its users.¹ Its versatility has allowed it to be utilized in various industries, including

^a Corresponding Author:
Teren Yedikian, MD
Rutgers New Jersey Medical School,
140 Bergen Street, ACC D1610, Newark, NJ, USA, 9739725350
ty210@njms.rutgers.edu

medicine. Within clinical practice, ChatGPT has been found to assist in generating accurate differential diagnoses, answering questions regarding medical conditions and queries, and creating patient letters, reports, and discharge summaries.²

Several surgical specialties have found the use of ChatGPT and other forms of artificial intelligence to be useful yet limited in clinical practice. A study done on the application of ChatGPT in clinical neurosurgical care stated that this tool was able to present information to patients regarding certain diagnoses; however, it could not complete more “human” aspects of care, such as assessment of emotional state and providing emotional support, such as using a physical exam to decipher diagnoses.³ However, this does present its potential viability to provide patient information sheets regarding a specific diagnosis. Within plastic and reconstructive surgery, ChatGPT’s limitations were similar to those found in neurosurgical studies. However, they found that the use of this tool did allow for increased productivity for plastic surgeons with regard to research and manuscript preparation, healthcare communication for patients, and drafting teaching sessions.⁴

There is little literature on the ability of ChatGPT to generate and present clinical information that patients can utilize regarding their medical conditions. However, within orthopaedics alone, the use of other forms of artificial intelligence has increased by a factor of 10, showing the viability of the use of a tool like ChatGPT in clinical practice.⁵ Studies have shown that artificial intelligence tools can be used to assist in the evaluation of diagnostic images, assessing injury risk, and other factors.⁵ Nevertheless, research on ChatGPT in orthopaedics has shown that on the current in-training exam for orthopaedic residents, it answered 190 out of 360 questions correctly, correlating to a post-graduate year one (PGY-1) level of knowledge.⁶ Thus, this tool may be useful in providing patient information rather than assisting clinicians in decision-making.

The objective of this study is to assess whether ChatGPT can generate patient data sheets that are comprehensive (also referred to as “readability”), understandable, and presented well for the average patient in the United States with respect to hand pathologies. The American Medical Association (AMA) recommends presenting medical information to patients at the sixth-grade reading level.⁷ We aim to compare these data sheets presented by ChatGPT to widely accepted patient information sheets from the American Academy of Orthopaedic Surgeons (AAOS), the American Association of Hand Surgery (AAHS), and the American Society for Surgery of the Hand (ASSH). We hypothesize that this tool will be able to present patient information in a more concise and readable manner compared to that of those written by each surgical society.

METHODS

Patient information sheets related to common hand pathologies were identified through the AAOS, AAHS, and ASSH websites in June 2023.⁸⁻¹⁰ All entries used were pathologies limited to the hand and wrist. ChatGPT version

3.5 was queried to generate patient information sheets on the same hand pathologies at the sixth-grade reading level. The following prompt was used to develop patient information sheets: “Create a patient information sheet about [hand pathology] for the 6th-grade reading level for the average United States (US) patient.” The articles from AAOS, AAHS, and ASSH, and text generated by ChatGPT, were copied in plain text and analyzed by a free-to-use readability website, WebFX.¹¹ Readability was calculated using the following readability scoring tests: Flesch Reading Ease, Automated Readability Index, Flesch Kincaid Grade Level, Gunning Fog Score, Simple Measure of Gobbledygook (SMOG) Index, and Coleman-Liau Index.

The Flesch Reading Ease is scored from 0 to 100, with higher scores correlating with easier readability.¹² A score of 0 to 30 and 30 to 50 corresponds with readable works that are scientific or academic in quality, respectively.¹³ A score of 80 to 90 corresponds with a 6th-grade reading level. The Automated Readability Index is the opposite in that a higher score has a lower readability.¹⁴ The remaining four scoring calculations are associated with the grade level needed to understand the text, where higher scores also have a lower readability.¹⁵⁻¹⁹

After filtering out articles in another language and pathologies not specific to the hand and wrist, we analyzed 28 articles from AAOS, 11 from AAHS, and 47 from ASSH. We asked ChatGPT to generate patient information sheets at the 6th-grade reading level using the same pathologies from each article in June 2023. The six readability scores were calculated for each article, and those generated from ChatGPT. Additionally, data on the number of sentences, words, complex words, percentage of complex words, average words per sentence, and average syllables per word were also collected.

A comparison of the readability between each article and those generated by ChatGPT was made using a paired two-tailed t-test. Statistical significance was defined as $P < 0.05$ for all tests.

RESULTS

Statistical analysis demonstrated that AAOS patient information sheets were associated with significantly higher Flesch Reading Ease, sentences, words, and complex words. Chat GPT was associated with significantly higher scores for Flesch Kincaid Grade Level, Gunning Fog Score, SMOG Index, Coleman Liau Index, Automated Readability Index, percent of complex words, and average words per sentence [Table 1].

AAHS patient information sheets had significantly higher Flesch Reading Ease and average words per sentence. In contrast, ChatGPT had significantly higher Coleman-Liau Index scores, percent of complex words, average words per syllable, and average syllables per word [Table 2].

ASSH sheets had significantly higher Flesch Reading Ease, number of sentences, and number of words. Chat GPT was associated with significantly higher scores for Flesch Kincaid Grade Level, Gunning Fog Score, SMOG Index,

Table 1. Mean Readability Scores of AAOS and ChatGPT Patient Information Sheets

	AAOS (mean) N = 28	ChatGPT (mean) N = 28	Δ (AAOS-ChatGpt)	P value
Flesch Kincaid Reading Ease	57.8±3.6	51.5±10.2	6.4±7.9	<.001
Flesch Kincaid Grade level	7.6±0.6	9.8±2.4	-2.3±2.0	<.001
Gunning Fog Score	8.6±1.0	12.5±2.8	-3.9±2.4	<.001
SMOG Index	7.4±0.5	9.6±1.9	-2.2±1.7	<.001
Coleman-Liau index	13.5±0.7	14.5±1.5	-1.0±1.2	<.001
Automated Readability	6.8±0.7	10.4±3.0	-3.6±2.6	<.001
Sentences	227.0±47.8	41±11.1	186.0±52.7	<.001
Words	2203.7±620.7	581.7±124.4	1622.0±656.2	<.001
Complex words	364.96±112.9	106.1±35.0	258.9±113.0	<.001
% Complex words	16.6±1.8	18.1±3.5	-1.5±2.6	0.005
Average words per sentence	9.6±1.0	15.1±4.3	-5.5±4.2	<.001
Average syllables per word	1.6±0.0	1.7±0.1	0.0±0.1	0.426

Table 2. Mean Readability Scores of AAHS and ChatGPT Patient Information Sheets

	AAHS (mean) N = 11	ChatGPT (mean) N = 11	Δ (AAHS-ChatGpt)	P value
Flesch Kincaid Reading Ease	59.9±9.6	53.0±4.9	6.9±9.5	0.037
Flesch Kincaid Grade level	8.7±1.9	8.9±1.5	-0.2±2.1	0.799
Gunning Fog Score	11.7±2.0	11.7±1.6	0.1±2.3	0.909
SMOG Index	8.7±1.4	8.9±1.1	-0.1±1.7	0.793
Coleman-Liau index	11.5±1.5	14.4±0.9	-2.9±1.8	<.001
Automated Readability	8.1±2.3	8.9±2.2	-0.8±2.7	0.371
Sentences	39.9±12.5	49.5±8.9	-9.6±18.8	0.119
Words	624.0±233.2	583.1±53.8	40.9±209.6	0.532
Complex words	94.6±45.9	111.5±14.3	-16.9±45.5	0.246
% Complex words	14.1±3.2	19.4±2.1	-5.3±3.3	0.001
Average words per sentence	15.4±2.5	12.3±3.7	3.2±4.3	0.035
Average syllables per word	1.6±0.1	1.7±0.0	-0.1±0.1	0.002

Coleman Liau Index, Automated Readability Index, percent of complex words, average words per sentence, and average syllables per word [Table 3].

DISCUSSION

ChatGPT is a novel AI system that has been taking the world by storm and can influence a patient’s perception of their pathology. Patients are more likely to be familiar with ChatGPT compared to patient information sheets from AAOS, AAHS, or ASSH. Therefore, it is imperative to be aware of the overall level of ease to read patient information sheets found on these academic sites and the ability to comprehend them. Based on the results, patient information sheets from AAOS and ASSH are significantly easier to read on most metrics. Those from AAHS are no different from those generated from ChatGPT, except for increased readability as measured by the Flesch Reading Ease and

Coleman-Liau Index. It should be noted that information sheets generated by ChatGPT have a significantly smaller number of words and complex words compared to those from AAOS and ASSH, suggesting that the information is more concise on ChatGPT.

Previous studies have shown that orthopedic patient material from the AAOS and ASSH has been difficult to read, with Flesch-Kincaid Reading Grade Level scores ranging from the 8th to 10th grade reading levels.^{15,20} Our study has shown that this material has been made easier to read over time, with a grade level score of around the 7th grade. Online educational material from institutions with both plastic and orthopedic surgery training programs is presented at the 11th-grade reading level, which is more difficult than those generated by ChatGPT and from the AAOS and ASSH.²¹

In terms of limitations and future directions, ChatGPT was instructed to create patient education material at the

Table 3. Mean Readability Scores of ASSH and ChatGPT Patient Information Sheets

	ASSH (mean) N = 47	ChatGPT (mean) N = 47	Δ (ASSH-ChatGpt)	P value
Flesch Kincaid Reading Ease	62.1±5.8	50.0±6.6	12.4±6.8	<.001
Flesch Kincaid Grade level	7.4±1.1	9.5±0.9	-2.1±1.2	<.001
Gunning Fog Score	9.3±1.5	11.7±1.1	-2.4±1.7	<.001
SMOG Index	7.6±0.8	9.2±0.7	-1.6±1.0	<.001
Coleman-Liau index	12.1±1.1	15.3±1.2	-3.2±1.1	<.001
Automated Readability	6.5±1.5	9.9±1.0	-3.4±1.5	<.001
Sentences	76.2±20.8	47.4±4.8	28.8±30.8	<.001
Words	886.3±470.1	603.6±46.3	282.8±469.9	<.001
Complex words	131.6±74.6	118.2±20.7	13.4±76.7	0.236
% Complex words	14.9±2.4	19.6±3.2	-4.7±2.8	<.001
Average words per sentence	11.3±2.0	12.8±1.0	-1.5±2.3	<.001
Average syllables per word	1.6±0.1	1.7±0.1	-0.13±0.1	<.001

6th-grade reading level. However, the Flesch-Kincaid Grade Level, on average, was at around an 8th to 9th grade reading level. Additionally, the Flesch Reading Ease of ChatGPT-generated material was in the 50s, which indicates that they are slightly easier to read than those that are academic in quality, instead of a score of 80 to 90, which corresponds with a 6th-grade reading level. One study has been able to train ChatGPT to convert orthopedic educational material to a lower grade level.²²

A metric to test medical validity should be measured between the two groups to compare the overall quality of the information. Other studies have used the Ensuring Quality Information for Patients tool; however, it has not been validated for AI chatbot use.²³ This study should be expanded to other subspecialties in orthopedics to investigate potential variations in pathologies.^{24,25} Furthermore, as ChatGPT is evolving as an AI, a repeat study in years to come should be conducted to see how much has changed since its inception.

CONCLUSION

While ChatGPT demonstrates the potential to generate concise patient education materials, its current outputs do not meet the recommended sixth-grade reading level set by the AMA. Compared to established hand surgery societies like AAOS and ASSH, ChatGPT-generated materials tend to be less readable across several standardized metrics, despite having fewer words and complex terms. These findings highlight both the promise and the limitations of using AI-generated content in patient education, particularly within hand surgery. As the technology continues to evolve, future efforts should focus on refining prompts, improving language model training, and incorporating validated tools to ensure both readability and medical accuracy. Expanding

this research to include other orthopedic subspecialties and repeating the study as AI tools improve will provide further insight into the utility and progression of ChatGPT in clinical patient communication.

.....
DECLARATION OF CONFLICT OF INTEREST

Teren Yedikian (No), Amanda Azer (No), Hershil Patel (No), Ariana Shaari (No), Brian Molokwu (No), Kush Modi (No), Irfan Ahmed (No), Michael Vosbikian (Honorarium for being the Section Head for Hand and Wrist for JBJS Clinical Classroom, Honorarium for being course faculty for the Medartis Pre-Hand Fellowship Primer, Editorial Board for ePlasty journal, and SurgiColl)

DECLARATION OF FUNDING

The authors received NO financial support for the preparation, research, authorship, and publication of this manuscript.

DECLARATION OF ETHICAL APPROVAL FOR STUDY

exempt from IRB as no human data were collected

DECLARATION OF INFORMED CONSENT

Not applicable

Submitted: May 21, 2025 EST. Accepted: August 10, 2025 EST.

Published: December 22, 2025 EST.

© All rights to the material are retained by the authors.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

REFERENCES

1. Lindner J. Must-know Chat GPT-4 statistics [recent analysis]. Gitnux. December 23, 2023. Accessed May 21, 2025. <https://gitnux.org/chat-gpt-statistics/>
2. Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. *J Med Internet Res*. 2023;25:e48568. doi:[10.2196/48568](https://doi.org/10.2196/48568)
3. Kuang YR, Zou MX, Niu HQ, Zheng BY, Zhang TL, Zheng BW. ChatGPT encounters multiple opportunities and challenges in neurosurgery. *Int J Surg*. 2023;109(10):2886-2891. doi:[10.1097/js9.0000000000000571](https://doi.org/10.1097/js9.0000000000000571)
4. Sharma SC, Ramchandani JP, Thakker A, Lahiri A. ChatGPT in plastic and reconstructive surgery. *Indian J Plast Surg*. 2023;56(4):320-325. doi:[10.1055/s-0043-1771514](https://doi.org/10.1055/s-0043-1771514)
5. 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(1):74. doi:[10.1186/s40634-023-00642-8](https://doi.org/10.1186/s40634-023-00642-8)
6. Kung JE, Marshall C, Gauthier C, Gonzalez TA, Jackson JB 3rd. Evaluating ChatGPT performance on the Orthopaedic In-Training Examination. *JB JS Open Access*. 2023;8(3). doi:[10.2106/jbjs.Oa.23.00056](https://doi.org/10.2106/jbjs.Oa.23.00056)
7. Weiss BD. *Health Literacy: A Manual for Clinicians*. American Medical Association Foundation and American Medical Association; 2003.
8. American Academy of Orthopaedic Surgeons. Diseases & conditions: hand & wrist. Accessed May 21, 2025. <https://orthoinfo.aaos.org/en/diseases--conditions/?bodyPart=HandWrist>
9. American Society for Surgery of the Hand. Hand conditions. Accessed May 21, 2025. https://www.assh.org/handcare/conditions#/+0/title_na_str/asc/
10. American Association for Hand Surgery. Accessed May 21, 2025. <https://handsurgery.org/>
11. WebFX. Readability test tool. Accessed May 21, 2025. <https://www.webfx.com/tools/read-able/>
12. Flesch R. A new readability yardstick. *J Appl Psychol*. 1948;32(3):221-233. doi:[10.1037/h0057532](https://doi.org/10.1037/h0057532)
13. Flesch R. *How to Write Plain English: A Book for Lawyers and Consumers.*; 1981:144.
14. Smith EA, Kincaid JP. Derivation and validation of the automated readability index for use with technical materials. *Human Factors: The Journal of Human Factors and Ergonomics Society*. 1970;12:457-564. doi:[10.1177/001872087001200505](https://doi.org/10.1177/001872087001200505)
15. Roberts H, Zhang D, Dyer GS. The readability of AAOS patient education materials: evaluating the progress since 2008. *J Bone Joint Surg Am*. 2016;98(17):e70. doi:[10.2106/jbjs.15.00658](https://doi.org/10.2106/jbjs.15.00658)
16. Kincaid JPF, Robert P Jr, Rogers RL, Chissom BS. Derivation of new readability formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) for Navy enlisted personnel. *Institute for Simulation and Training*. 1975;56. doi:[10.21236/ADA006655](https://doi.org/10.21236/ADA006655)
17. McLaughlin GH. SMOG grading: a new readability formula. *Journal of Reading*. 1969;12(8):639-646.
18. Coleman M, Liau TL. A computer readability formula designed for machine scoring. *Journal of Applied Psychology*. 1975;60:283-284. doi:[10.1037/h0076540](https://doi.org/10.1037/h0076540)
19. Gunning R. *The Technique of Clear Writing*. McGraw-Hill; 1952.
20. Wang SW, Capo JT, Orillaza N. Readability and comprehensibility of patient education material in hand-related web sites. *J Hand Surg Am*. 2009;34(7):1308-1315. doi:[10.1016/j.jhsa.2009.04.008](https://doi.org/10.1016/j.jhsa.2009.04.008)
21. Cook JA, Sasor SE, Tholpady SS, Momeni A, Chu MW. Hand surgery resources exceed American health literacy. *Hand (N Y)*. 2018;13(5):547-551. doi:[10.1177/1558944717725384](https://doi.org/10.1177/1558944717725384)
22. Kirchner GJ, Kim RY, Weddle JB, Bible JE. Can artificial intelligence improve the readability of patient education materials? *Clin Orthop Relat Res*. 2023;481(11):2260-2267. doi:[10.1097/corr.0000000000002668](https://doi.org/10.1097/corr.0000000000002668)
23. Walker HL, Ghani S, Kuemmerli C, et al. Reliability of medical information provided by ChatGPT: assessment against clinical guidelines and patient information quality instrument. *J Med Internet Res*. 2023;25:e47479. doi:[10.2196/47479](https://doi.org/10.2196/47479)
24. Parsa A, Ebrahimzadeh MH. ChatGPT in Medicine; a Disruptive Innovation or Just One Step Forward? *Arch Bone Jt Surg*. 2023;11(4):225-226. doi:[10.22038/abjs.2023.22042](https://doi.org/10.22038/abjs.2023.22042)

25. Bruey K, Kachooei AR. Applications, Implications, and Drawbacks of Artificial Intelligence in Medical Publications. *Arch Bone Jt Surg.* 2025;13(1):1-3.
doi:[10.22038/ABJS.2024.82343.3751](https://doi.org/10.22038/ABJS.2024.82343.3751)