

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 Yedikian¹, Hershil Patel¹, Ariana Shaari¹, Brian Molokwu¹, Kush Modi¹, Irfan Ahmed¹, Michael Vosbikian¹
¹Rutgers New Jersey Medical School, Newark, NJ
 Ty210@njms.rutgers.edu

Disclosures: Teren Yedikian (N), Hershil Patel (N), Ariana Shaari (N), Brian Molokwu (N), Kush Modi (N), Irfan Ahmed (N), Michael Vosbikian (Honorarium for being the Section Head for Hand and Wrist for JBSJ Clinical Classroom, Honorarium for being course faculty for the Medartis Pre-Hand Fellowship Primer, Editorial Board for ePlasty and SurgiColl journals)

INTRODUCTION: Artificial intelligence (AI) language models such as Chat GPT have experienced a recent rise in popularity. Chat GPT, launched in fall 2022, is a conversational tool that possesses the ability to compose, analyze, and present information to its users. Its versatility has allowed it to be utilized in various industries including medicine. There is little literature on the ability of Chat GPT to generate and present clinical information that patients can utilize regarding their medical conditions. The objective of this comparative study was to explore whether Chat GPT can generate patient information sheets on common hand pathologies for the average US patient. Previous literature has demonstrated that the average US patient reads at the sixth grade reading level. The comprehensibility (dubbed “readability”) of Chat GPT 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. Patient information sheets on various hand pathologies that occur in US patients were compared in terms of understandability and presentation.

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. Chat GPT was utilized to generate patient information sheets on the same hand pathologies at the sixth grade reading level. WebFx was utilized to calculate readability (Flesch-Kincaid Reading Ease, Flesch Kincaid Grade Level, Gunning Fog Score, SMOG Index, Coleman LIAU Index, Automated Readability Index) scores for the AAOS, AAHS, ASSH and Chat GPT-generated patient sheets. Statistical analysis was determined using a paired two-tailed t-test; moreover, an ANOVA was performed to compare the four groups. Statistical significance was defined as $p < 0.05$.

RESULTS SECTION: Statistical analysis demonstrated that AAOS patient information sheets were associated with significantly higher Flesch-Kincaid 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-Kincaid Reading Ease and average words per sentences, whereas 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-Kincaid 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, 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. Since patients are more likely to be more aware of ChatGPT compared to patient information sheets from AAOS, AAHS, or ASSH, it is imperative to know the overall level of ease to read patient information sheets and understand it. 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 than those generated from ChatGPT except for being easier to read per the Flesch-Kincaid Reading Ease and Coleman LIAU Index. It should be noted that information sheets generated by ChatGPT have a significantly fewer number of words and complex words compared to those from AAOS and ASSH, suggesting that information is more concise in ChatGPT.

In terms of limitations and future directions, a metric to test for medical validity should be measured between the two groups to compare the overall quality of the information. This study should be expanded to other subspecialties in Orthopedics to see if there is any variation between pathologies. 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 2023.

SIGNIFICANCE/CLINICAL RELEVANCE: (1-2 sentences): Orthopedic Surgeons can provide patients with information sheets on common hand pathologies from the AAOS and ASSH, knowing that they are easier to read compared to those generated by ChatGPT.

IMAGES AND TABLES:

Table 1: Paired t-test Comparing Readability Between 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 word 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: Paired t-test Comparing Readability Between 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 word 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

Table 3: Paired t-test Comparing Readability Between 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 word 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