

Evaluation of Patient Education Materials from Large-Language Artificial Intelligence Models on Carpal Tunnel Release: Are Patients Learning?

Brett J Croen¹, Mohammed S. Abdullah¹, Sarah Rapaport¹, Ellis Berns¹, Alexander Hahn², Caitlin Barrett¹, Andrew Sobel¹

¹Department of Orthopaedic Surgery, The Hospital of the University of Pennsylvania, ²Department of Orthopaedic Surgery, UConn MSK Institute
Brett.Croen@penmedicine.upenn.edu

Disclosures: BJC (N), MSA (N), SR (N), EB(N), AH (N), CB (N), AS (N)

INTRODUCTION: Abundant healthcare information on Carpal Tunnel Release (CTR) is available on the internet, rendering it a crucial tool for patients aiming to enhance their comprehension of diseases and explore various treatment options. Past studies have indicated that as much as 80% of internet users will search for healthcare-related information [1], underscoring the internet's effectiveness in investigating illnesses, medical therapies, procedures, and other health-related details. Many patients turn to the internet to shape their medical choices, as evidenced by a 2006 survey in which 55% of participants acknowledged that online healthcare information heightened their access to medical knowledge, and 47% noted a noteworthy impact on their medical decision-making. Furthermore, with the commercialization of Artificial intelligence that utilizes large language models trained on diverse text sources, many patients now have started utilizing other resources aside from traditional search engines for routine health related questions [2]. ChatGPT (Open AI, San Francisco, CA), the most utilized AI software publicly available for use, has the potential to be a useful patient aid. It is imperative to compare the sources and characteristics of responses from ChatGPT and Google when querying with common questions asked by patients regarding CTR surgery to see what patients are learning about CTR. Currently, The American Medical Association (AMA) and National Institutes of Health (NIH) recommend that patient education materials be written at no higher than a sixth-grade reading level with previous studies on the topic showing most patient materials on CTS exceed both the readability limits recommended by the AMA and NIH, and the average reading ability of most US adults [3]. This study seeks to assess the accuracy and readability of CTR information provided by the AI software ChatGPT. Utilizing standardized scoring methods, we analyze the quality of information and document AI-generated responses.

METHODS: A Google Web Search (Alphabet, San Jose, CA) was performed using a newly installed Google Chrome browser. Before the search, the browser was confirmed to be clean to minimize the effects of search algorithms. This was defined as one that cleared files, sponsored sites/advertisements, browsing history, cookies, and caching images/files. The term entered was “carpal tunnel release surgery,” and the first ten Frequently Asked Questions (FAQs) were recorded (**Table 1**). FAQ questions were obtained by clicking on the “People Also Ask” header under the search bar. The inclusion criteria included any question with the term “carpal tunnel” “carpal tunnel surgery” and “carpal tunnel release.” Exclusion criteria included any repeat questions or questions unrelated to CTR. The questions were categorized into question topics: fact, policy, and value using the Rothwell Classification System. The website sources were categorized into commercial, academic, medical practice, single surgeon practice, government, and social media using previously established categories from the literature. The first inputted text in ChatGPT was “Record the ten most popular questions related to Carpal Tunnel Release Surgery.” The top ten questions were recorded and compared to those from Google. Similar questions are represented in bold font. To properly compare answer content, ChatGPT was then asked to individually answer each of the ten Google FAQs and provide sources. Answer content was analyzed using word count as well as three different readability analyses: Flesh Reading Ease score, Gunning-Fog index, and Dale Call Readability formula. The significance level was set at $p < 0.05$. Statistical analysis was performed via Independent t-test (Prism, Graphpad)

RESULTS: Most Frequently Asked Questions on Google vs ChatGPT: Of the ten most frequently asked questions on Google and ChatGPT, 40% were the same between the two search engines. The Google FAQs topics were 40% fact, 30% value, and 30% policy while the ChatGPT topics were 40% policy, 30% value, and 30% fact. **Answers to Google's Ten Most Common FAQs:** Google answers to FAQs were from one source per question: commercial (40%), academic (40%), government (10%), and single-surgeon practice (10%) websites. With ChatGPT, each answer was constructed from two different sources per question and the sources were all from academic websites including The American Academy of Orthopaedic Surgeons (AAOS), American Society for Surgery of the Hand (ASSH), Mayo Clinic, and Cleveland Clinic. A comparison between the answer content revealed that the ChatGPT answers (289 ± 80 words) were significantly longer than Google's answers (43 ± 7 words) (p -value < 0.0001). The three questions that had numerical answers 1. How long is recovery from carpal tunnel surgery; 2. How long does carpal tunnel release surgery take; 3. When is it too late for carpal tunnel surgery?) revealed similar answers.

Readability Analyses: The Flesh Reading Ease Score algorithm calculated Google to be 45.54 ± 17.83 and ChatGPT at 28.9 ± 8.2 ($p = 0.02$). The Gunning Fog index showed Google's grade level to be 14.95 ± 3.77 and ChatGPT to be at 18.22 ± 1.73 ($p = 0.02$) and the Dale-Chall's Readability Formula, determined Google's readability grade level to be 8.73 ± 2.93 and ChatGPT's to be 10.98 ± 0.71 ($p = 0.03$). Finally, all ChatGPT answers ended with some variation of an instruction to consult your surgeon or healthcare provider in order to get patient-specific and more accurate answers.

DISCUSSION: Our findings reveal that there was a 40% overlap in the most common CTR Related questions between ChatGPT and Google; that ChatGPT and Google had similar ratios of fact, value, and policy-based questions; and that ChatGPT provided longer, more detailed responses than Google. In the provided answers, ChatGPT utilized two academic sources, whereas Google utilized only one source, which was most commonly either commercial or academic. Despite different sources, the three questions that had numerical-based answers revealed comparable timelines. On readability analysis, Google's responses were easier to read than ChatGPT's when measured by the Flesh Reading Ease Score and were at a lower grade level than ChatGPT's when measured by the Gunning Fog Index and Dale-Chall Readability Formula. Both Google and ChatGPT have patient education material written at an educational level above the NIH and AMA recommendations; as a result, average American patients may not be able to fully understand them.

SIGNIFICANCE: Reevaluating patient education materials in CTR could enhance patient comprehension of their health status and treatment choices, leading to more effective and efficient physician-patient interactions.

REFERENCES: [1] Fox + Internet and American Life Project, 2011, [2] Rice + PEW Research Center, 2006, [3] Sheriff + Hand, 2017

Table 1: Ten Most Frequently Asked Questions According to ChatGPT and Google	
ChatGPT	Google
What is carpal tunnel release surgery? (Fact)	How long is recovery from carpal tunnel surgery? (Fact)
How long does carpal tunnel release surgery take? (Fact)	How painful is carpal tunnel surgery? (Value)
Is carpal tunnel release surgery painful? (Value)	Is carpal tunnel release a major surgery? (Value)
What are the risks and complications of carpal tunnel release surgery? (Policy)	How long does carpal tunnel release surgery take? (Fact)
How long is the recovery period after carpal tunnel release surgery? (Fact)	What to expect immediately after carpal tunnel surgery? (Fact)
Will I need physical therapy after carpal tunnel release surgery? (Policy)	When is it too late for carpal tunnel surgery? (Policy)
How effective is carpal tunnel release surgery in relieving symptoms? (Value)	What is worse after carpal tunnel surgery? (Policy)
Can carpal tunnel syndrome come back after surgery? (Value)	What are the cons of carpal tunnel surgery? (Policy)
Are there any alternative treatments to carpal tunnel release surgery? (Policy)	Why does carpal tunnel surgery hurt so bad? (Value)
What is the cost of carpal tunnel release surgery? (Fact)	Should I wear a wrist brace after carpal tunnel surgery? (Policy)

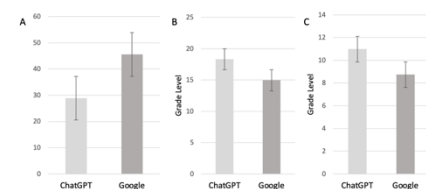


Figure 1: Readability Analysis (A) Flesh Reading Ease score ranges 0-100 (0 = difficult to read, understood by university graduates; 100 = easy to read, understood by an average 11-year-old). (B) Gunning-Fog Grade Index's and (C) Dale-Chall Readability Grade Index's value corresponds grade level