

Comparing ChatGPT and Online Information for Common Patient Questions in Orthopaedic Trauma

*Hamza Raja; **Muhammad Jehad Abbas, MD**; Hamza Kanchwala; Idris Nagarwala; Lindsay M Maier, MD; Stuart Trent Guthrie, MD; William Michael Hakeos, MD; Joseph John Hoegler, MD*

Purpose: Patients have been utilizing the internet as a source of medical and surgical information at an increasing rate. Prior literature has cautioned on the quality and readability of medical information encountered online. Recent advancements in artificial intelligence (AI) and large language models such as ChatGPT (version 3.5, OpenAI) caused patients to utilize this software for medical information. The purpose of this study was to investigate the quality and readability of ChatGPT responses to common patient questions in orthopaedic trauma compared to current reputable online information.

Methods: A series of 20 common patient questions and responses (control) were taken from the OTA patient frequently asked questions (FAQs) website. These questions were asked to ChatGPT on November 13, 2023, and responses were recorded (AI). The original questions, the blinded control responses, and the blinded ChatGPT responses were provided to 4 fellowship-trained orthopaedic trauma surgeons to grade for 3 components: preference, medical accuracy, and appropriateness. Accuracy and appropriateness were graded using a Likert scale. Additionally, each response (AI and control) was graded for readability using the Flesch-Kincaid Grade Level.

Results: Surgeons identified no significant preference between AI and control responses ($P > 0.05$). ChatGPT responses had a significantly greater mean accuracy than control responses (18.0 ± 2.0 vs 15.6 ± 2.58 , respectively; $P < 0.05$). Additionally, ChatGPT more appropriately answered questions compared to control responses (17.6 ± 2.1 vs 15.2 ± 3.8 , respectively; $P < 0.05$). ChatGPT provided answers at a higher reading level (AI: 13.9 years \pm 2.2, control: 8.3 years \pm 1.4; $P < 0.05$) and with significantly more words than the control response (AI: 179.15 words \pm 66.4, control: 119.6 words \pm 63; $P < 0.05$).

Conclusion: ChatGPT is a suitable option for patients looking online to learn answers to common questions in orthopaedic trauma. Responses by ChatGPT were found to be accurate and appropriate when compared to a reputable control, offering a valuable alternative to information available online. ChatGPT responses were written at a significantly higher reading level. As with online searches, patients should express caution with information presented and defer to surgeon counseling when making treatment decisions.