Can ChatGPT-4 Be Used to Answer Patient Questions Concerning the Latarjet Procedure for Anterior Shoulder Instability?

Amy Lu[1], Michael Robert Mazzucco, Ayoosh Pareek, Michael Fu, David M Dines[2], Russell F Warren[3], Lawrence V Gulotta[4], Joshua S Dines[2], Kyle Kunze[5]
[1]Weill Cornell Medicine, [2]Hospital For Special Surgery, [3]Hosp for Special Surgery, [4]Hosp for Special Surg-Cornell, [5]Orthopaedic Surgery, Hospital for Special Surgery

INTRODUCTION: Surgeons and advanced practitioners are continuously challenged with patient inquiries concerning their musculoskeletal diagnosis or postoperative instructions. Responding to queries may be time sensitive depending on clinical diagnosis and also have implications for patient satisfaction. The purpose of the current study was to assess the propensity for ChatGPT-4, an automated Chatbot powered by artificial intelligence (AI), to provide medical information concerning the Latarjet procedure for patients with anterior shoulder instability.

METHODS: Using previously validated methods, a Google search was first performed using the query "Latarjet." Subsequently, the top ten frequently asked questions (FAQs) and associated sources were extracted. ChatGPT-4 was then prompted to perform a Google search using the query "Latarjet" and to provide the top ten FAQs concerning the procedure. Google was again prompted for the top ten FAQs requiring discrete-numeric answers, and these FAQs were asked to ChatGPT-4 to comparatively assess validity and concordance. Comparisons of content and sources utilized by ChatGPT-4 and Google were performed using Fischer's exact or Chi-squared tests where appropriate, with a p-value <0.05 indicating statistical significance.

RESULTS:

Substantial or identical overlap for answers to FAQs requiring numeric responses were found in 90% of queries, with ChatGPT-4 deriving information for answers from academic sources in all instances (**Figure 1**). This was significantly different from Google (p<0.001), which used only 30% academic sources, with the remaining consisting of information found on surgeon personal websites (50%) and larger medical practices (20%). For general FAQs, 40% of FAQs were found to be identical when comparing ChatGPT-4 and Google. ChatGPT-4 again used 100% academic resources, while Google used 60% academic resources, 20% surgeon personal websites, and 20% medical practices. The most common question category for both ChatGPT-4 and Google was technical details (40%, **Figure 2**). ChatGPT-4 also presented information concerning risks/complications (30%), recovery timeline (20%), and evaluation of surgery (10%).

DISCUSSION AND CONCLUSION: In a replicated patient scenario where information concerning the Latarjet procedure was sought, ChatGPT-4 demonstrated the capability to provide a broad range of clinically relevant questions and answers. ChatGPT-4 derived information from academic sources in 100% of cases for both general FAQs and those requiring numeric responses, while the sources used by Google were heterogeneous. Pending further training and refinement, ChatGPT-4 may be utilized as an adjunct for shoulder surgeons and ancillary staff to decrease the burden of patient inquiries.

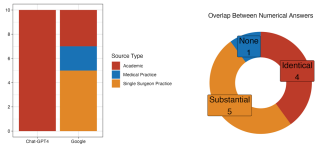

Figure 1. Sources and overlap for frequently asked questions with discrete-numeric answers.

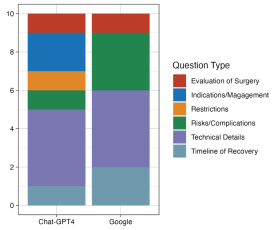

Figure 2. Question-type categorizations for open-ended frequently asked questions.