

ChatGPT Answers Patient Focused Questions in Cervical Spine Surgery

Tejas Subramanian, Kasra Araghi, Branden Rafael Sosa, Troy B Amen, Omri Maayan, Pratyush Shahi¹, James E Dowdell, Sheeraz Qureshi², Sravisht Iyer³

¹Hospital For Special Surgery, ²Minimally Invasive Spine Surgery, ³Hospital for Special Surgery

INTRODUCTION:

Artificial intelligence (AI) has the potential to revolutionize how medicine is practiced. Across medicine, from radiology to drug development, we are seeing remarkable growth in the field. One such way AI has the potential to impact clinical medicine is through patient education. AI chatbots such as ChatGPT can provide answers to a seemingly infinite number of questions. While patients may turn to these recourses when making orthopaedic decisions, their accuracy and utility have not yet been thoroughly assessed.

METHODS:

We compiled 9 frequently asked questions cervical spine surgeons receive in clinic to test ChatGPT's version 3.5 ability to answer a nuanced topic. Questions were inputted one at a time into the ChatGPT user interface. Questions included were: 1) What are the different types of cervical spine surgery? 2) What are the indications for cervical spine surgery? 3) What is the difference between an anterior cervical discectomy and fusion (ACDF) and a cervical disc replacement (CDR)? 4) What are the benefits and drawbacks of ACDF and CDR? 5) Will my neck pain improve after ACDF or CDR? 6) Will my arm pain improve after ACDF or CDR? 7) Is there a chance I will not improve or get worse after cervical spine surgery? 8) What are the chances I will have swallowing difficulties following cervical spine surgery? 9) How long is the recovery following cervical spine surgery? Responses were reviewed by two independent reviewers for the accuracy of information presented (0-5 points), appropriateness in giving a specific answer to specific question asked (0-3 points), and readability for a layperson (0-2 points). Responses were also qualitatively assessed for potential limitations of the chatbot.

RESULTS: On average, ChatGPT's responses scored an 8.1/10. Accuracy was rated on average a 3.9/5. Appropriateness was 2.2/3. Both readers rated ChatGPT a 2/2 for readability for all 9 questions.

DISCUSSION AND CONCLUSION:

As chatbots such as ChatGPT become increasingly popular, they are sure to become another resource for patients considering orthopaedic surgery. Assessing ChatGPT's capacity to address cervical spine related patient questions revealed several strengths and limitations. On average, ChatGPT scored a 3.9 out of 5 points for accuracy. This is an impressive demonstration of the data the chatbot has access to as these specific spine surgery questions are not easily answered. However, more significantly, both observers noted that despite ChatGPT's accuracy, it failed to provide a comprehensive response, often omitting important considerations that compromise the utility of the response. Appropriateness and readability were the two significant strong points for ChatGPT. Often times if a patient searches for a specific question, it can be 1) difficult to understand research literature or 2) not available on websites such as WebMD. ChatGPT was able to provide answers that answer the specific question asked in a way that most can understand. And because of this, it becomes a very powerful tool for the dissemination of information to patients looking to understand their pathologies and surgical options. Even more importantly, ChatGPT while would provide a response, it is always prefaced with a statement regarding consulting a spine professional for specific medical advice. We believe that patients will increasingly utilize ChatGPT as a resource for informing their medical decision making. Since the information may not always be fully accurate or inclusive, surgeons should know patients will be using it as a tool and should be aware of its limitations.