

ChatGPT as a Source of Patient Information for Lumbar Spinal Fusion and Laminectomy: A Comparative Analysis Against Google Web Search

Mitchell Kai-Sern Ng¹, Patrick Nian², Matthew Magruder³, Ian Wellington, Brian Ford, Nikhil Vasireddi, Jack Choueka⁴, Afshin Razi³

¹Orthopaedic Surgery, Maimonides Medical Center, ²Department of Orthopaedic Surgery, SUNY Downstate, ³Maimonides Medical Center, ⁴Maimonides Bone and Joint Center

INTRODUCTION: ChatGPT, an artificial intelligence chatbot with seemingly unlimited functionality, may present an alternative to a Google web search for patients seeking medical information. The objective of this study was to assess the utility of ChatGPT in providing patient information for lumbar spinal fusion and lumbar laminectomy in comparison with the Google search engine, with respect to: (1) the top ten most frequently asked questions (FAQs) associated with these procedures by question type/topic; (2) answers provided for shared FAQs; and (3) replicated responses in ChatGPT from FAQs with a numerical response on Google web search.

METHODS:

The search terms “lumbar spinal fusion” and “lumbar laminectomy” were individually inputted into a Google web search and the first 10 FAQs were extracted. ChatGPT was then utilized to extract its own list of 10 FAQs using the statement: “Perform a google search with the search term ‘lumbar spinal fusion’ and record the 10 most FAQs related to the search term”, repeated with ‘lumbar laminectomy’ as the second search term. Shared questions were inputted into ChatGPT for a response, and responses between two platforms were compared in terms of length/readability, using Flesch Reading Ease score and Flesch-Kincaid Grade Level.

RESULTS: 2 of 10 (20%) questions for both lumbar spinal fusion and lumbar laminectomy were asked similarly between ChatGPT and Google web search. Compared to the Google web search, ChatGPT’s responses to these questions were a mean length of 340.0 words (vs. 159.3), an average Flesch Reading Ease score of 34.0 (vs. 58.2), and an average Flesch-Kincaid grade level of 11.6 (vs. 8.8). Subjectively, each response concluded with a recommendation for the user to discuss further with a healthcare provider.

DISCUSSION AND CONCLUSION:

FAQs and responses regarding lumbar spinal fusion and lumbar laminectomy were highly variable between ChatGPT and a Google web search. While ChatGPT may be able to produce relatively accurate responses in select questions, its role remains as a supplement or starting point to a consultation with a physician, not as a replacement. Its utility as a resource for patient information must be taken with caution until functionality of its response production can be verified/validated.