Comparative Analysis of Clinical Relevance and Accuracy in AI-Assisted Patient Consultations on Ankle and Clavicle Fracture Surgeries

Daniel E Pereira, Helena Barber, Carrie Reaver, Anna Noel Miller

INTRODUCTION:

It is becoming increasingly important to evaluate the effectiveness of large language models (LLMs) and query-assisted platforms like Google and ChatGPT in providing clinically relevant and accurate information to patient-initiated inquiries. Limited studies have characterized the performance of these platforms on semi-elective orthopedic trauma surgeries. This study evaluates the function of both interactive online models on frequently queried patient searches on ankle and clavicle fracture operations.

METHODS:

An inquiry set consisting of ten prevalent patient questions for each fracture type was extracted using Google and ChatGPT. Responses, alongside their source attributions, were curated for clinical relevance and accuracy. This assessment was conducted in a double-blind fashion. Grading was completed by two researchers with academic and clinical experience in orthopedic trauma (D.E.P, H.F.B), with oversight and validation provided by a board-certified orthopedic trauma surgeon (A.N.M). Descriptive and comparative statistics were used to understand the dataset.

RESULTS:

ChatGPT performed better than Google in terms of clinical relevance ($p=0.001$) and accuracy ($p=0.004$) in ankle fracture patient queries. Clavicle fracture queries showed a significant difference favoring ChatGPT for accuracy ($p=0.04$), though relevance did not reach statistical significance ($p=0.06$). When answer ground-truth was analyzed, GPT outperformed Google by providing more academic sources ($p<0.05$).

DISCUSSION AND CONCLUSION:

Large Language Models outperformed traditionally used online platforms in delivering clinically precise and contextually relevant information on semi-elective, common orthopedic trauma surgeries. The ability of LLMs to synthesize responses from credible medical sources significantly diminishes the variability and potential inaccuracies inherent in conventional web searches. These insights strongly suggest that LLMs could play a pivotal role in enhancing patient engagement and comprehension in trauma care, meriting further exploration of their integration within clinical frameworks.