

Can Generative AI Resolve Patient Queries? An Analysis of AI Responses to Common Patient Queries after Ankle Fractures

Vivek Anand Pisharody, Aaron Daniel Therien, Grayson H Nour, Suman Medda, Blake Joseph Schultz, Andrew Adamczyk, Peiyu Li, Chuan Hong, Christian Alexander Pean

INTRODUCTION: Responding to patient questions after treatment can require significant time and staff resources. Generative artificial intelligence (AI) has received significant interest as a potential modality for triaging and resolving patient queries. However, the accuracy and safety of AI responses in the setting of orthopedic trauma has not been fully evaluated. In this study, we evaluate whether a generative AI model trained on relevant orthopedic literature can appropriately respond to common patient questions after initial management of an ankle fracture.

METHODS: Papers relevant to operative and non-operative management of ankle fractures were identified and used to train an AI agent in a commercially-available AI platform. A series of hypothetical queries representative of common patient concerns after initial operative or nonoperative management of an ankle fracture were written. Queries were submitted to the AI agent. All responses were rated for safety, clinical accuracy, and understandability on a five-point Likert scale by orthopedic trauma faculty.

RESULTS: Across all reviewers, a rating of 1/5 (excellent) or 2/5 (very good) was achieved by 92.9% of queries when evaluated for safety and 91.1% for understandability, but only 88.1% for accuracy. Mean medical safety scores per rater were high (range 1 to 1.39). Mean accuracy score varied between raters from 1.04 to 1.61. All reviewers rated understandability of responses highly (mean 1.02 to 1.28 per rater). However, there were significant differences in ratings of individual queries across scoring categories, with an intraclass correlation (ICC) of 0.17 for safety, 0.27 for accuracy, and 0.12 for understandability, suggesting significant interrater variability.

DISCUSSION AND CONCLUSION: The use of AI to address patient concerns is a rapidly expanding area of clinical interest. All reviewers in this study reported high safety, accuracy, and understandability scores overall. However, individual queries received differing scores from different raters. These results suggest that AI tools can be cautiously employed to reduce the administrative burdens of responding to patient queries.