Trauma Surgeon Evaluation of a Large Language Model's Postoperative Guidance for Tibial Plateau Fracture Patients

Introduction: Patients increasingly rely on commercially available large language models (LLMs) as for musculoskeletal health information, especially during periods when direct communication with their care team may be limited. However, the accuracy, safety, and clinical appropriateness of LLM-generated postoperative advice for orthopedic trauma patients remains unclear. The objective of this study was to evaluate trauma surgeons' assessments of the accuracy, safety, and comfort with answers provided by Google Gemini to patient's frequently asked questions after tibial plateau fractures at the knee.

Methods: An expert panel consisting of trauma surgeons and advanced practice providers identified the most commonly asked questions in the post-operative period following tibial plateau fracture fixation surgery. Each question was submitted individually to public Google Gemini Flash 2.5 in separate sessions. Two fellowship-trained trauma surgeons were asked to rate the accuracy and safety of the answer and their comfort level if their patient received the output.

Results: Thirty-two questions were evaluated. Accuracy ratings were consistently high, with 87.5% of responses receiving the maximum accuracy score (>80%) and the remaining 12.5% rated as 60-80% accurate. Safety ratings were similarly favorable, with 96.9% of responses rates as safe and 3.1% rates as questionably safe. Surgeons reported being very comfortable with 78.1% of responses and somewhat comfortable with the remaining 21.9%.

Conclusions: While Gemini generated generally accurate and safe postoperative advice for tibial plateau fracture patients, incomplete surgeon comfort indicates important nuance may be missing. Patients should view LLMs as supportive tools but still contact their care team for individualized guidance, and surgeons may need to clarify common gaps in LLM-generated guidance.