Evaluating AI Language Models for Patient Queries on Total Knee Replacement (TKR)

Anesu Karen Murambadoro
The University of Texas Rio Grande Valley School of Medicine, anesu.murambadoro01@utrgv.edu

Victoria Elizondo
The University of Texas Rio Grande Valley School of Medicine, victoria.elizondo03@utrgv.edu

Brianna Guillen
The University of Texas Rio Grande Valley School of Medicine, brianna.guillen01@utrgv.edu

Matthew Hnatow
The University of Texas Rio Grande Valley School of Medicine, matthew.hnatow@utrgv.edu

Michael Sander
The University of Texas Rio Grande Valley School of Medicine, michael.sander@utrgv.edu

Title: Evaluating AI Language Models for Patient Queries on Total Knee Replacement (TKR)

Authors/Presenters: Murambadoro AK, Elizondo V, Guillen B, Hnatow M, Sander M

Introduction:
Within the past few years, large language models (LLMs) (ChatGPT, LLaMa 3, Microsoft Copilot) have increasingly become a resource that patients engage with to learn about health care procedures, including total knee replacement (TKR). Previous studies have analyzed the efficacy of large language models in providing accurate and relevant responses to questions about various procedures. Our study aims to evaluate the clarity, validity, and understandability of LLMs to patient questions about total knee replacement and assess the consistency of these models and their effectiveness in providing accurate, valid, and guideline-adherent information to patients.

Methods: We selected 30 frequently asked questions for TKR in five categories: preoperative concerns, operative details, postoperative recovery, complications and lifestyle changes post-surgery. The questions were posed to AI models including LLaMA 3, ChatGPT-4.0, Microsoft Copilot, Google's Bard and Perplexity. Clinical orthopedic surgeons specializing in TKR assessed LLM responses on a Likert scale to grade their clarity, validity and understandability.

Results (pending): Data analysis is ongoing. However, preliminary results are expected by August 10 and will be presented at the symposium.

Conclusions (pending): This study is expected to demonstrate the effectiveness of AI language models in providing clear, valid, and understandable information about total knee replacement to patients. The results will help create a better understanding of how these models can be integrated into patient education and support, highlighting areas for improvement and further study to enhance their reliability and usefulness in clinical practice.