# Can ChatGPT Replace Orthopedic Triage Nurses?

Stephen W. Chenard, MSc[1]; Aleksander P. Mika, MD[1]; Stephen M. Engstrom, MD, MBA[1]; Gregory G. Polkowski, II, MD, MSCR[1]; Jacob M. Wilson, MD[1]; J Ryan Martin, MD[1]

[1]Department of Orthopaedic Surgery, Vanderbilt University Medical Center, Nashville, TN
stephen.w.chenard.1@vumc.org

**Disclosures:** S.W. Chenard: None. A.P. Mika: None. S.M. Engstrom: 3B; Enovis, LinkBio. 9; AAKHS - Advocacy Committee, AAOS - Coding Coverage and Reimbursement Committee. G.G. Polkowski: 1; Enovis. 3B; Enovis. 9; AAHKS. J.M. Wilson: 3B; AccuPREDICT. J.R. Martin: 3B; DePuy.

**INTRODUCTION:** Nurse triage remains a considerable financial and time burden following primary total joint arthroplasty (TJA). Furthermore, reimbursement continues to decrease leaving fewer resources available to deliver important perioperative care. Triage methods that are equally effective and free would be ideal. Therefore, the purpose of this study was to see if ChatGPT could serve as a safe and effective replacement for triage nurses in responding to frequently asked questions (FAQs) following joint replacement surgery.

**METHODS:** We compiled a list of ten FAQs from post-operative arthroplasty patients and presented them to orthopedic triage nurses (n=3) and ChatGPT (GPT-3.5 Free Research Version Aug 23) (Table 1). The responses generated by ChatGPT were based on distinct conversations without any follow-up or repetition. The responses were rated by three blinded fellowship-trained arthroplasty surgeons using Likert scales ranging from 1 to 5 for safety and completeness. The mean scores for each group were calculated across all surgeons and FAQs and are presented as mean ± standard deviation. Paired T-tests were used to compare triage nurse and ChatGPT responses.

**RESULTS:** There was no significant difference in average safety scores between ChatGPT responses (4.233 ± 0.473) and those of triage nurses (4.278 ± 0.382) (p=0.68) (Fig 1). Similarly, there was no significant difference in completeness scores between ChatGPT responses (3.933 ± 0.783) and those of triage nurses (4.044 ± 0.588) (p=0.67) (Fig 2). Furthermore, when examined on the individual prompt level, there were no significant differences in safety or completeness between ChatGPT and triage nurses' responses (all p-values >0.05).

**DISCUSSION**: As reimbursement for primary TJA continues to decrease, utilizing novel methods for providing cost-effective clinical care will become paramount. Nurse triage is an important component of perioperative care that is associated with substantial costs. Our study demonstrates that ChatGPT can provide safe and complete triage responses instantaneously. Accordingly, we feel that artificial intelligence tools should be developed and trained in a domain-specific fashion for patient triage purposes and particular promise may lie in electronic medical record integration.

**SIGNIFICANCE:** While preliminary, this work highlights the potential impact of artificial intelligence in responding to post-operative questions from orthopedic patients and will likely be applicable to other surgical sub-specialties as well.

| Table 1. List of FAQ prompts provided to triage nurses and ChatGPT |
|---|
| 1) "How much walking can I do?" |
| 2) "My knee hurts. What should I do?" |
| 3) "My incision is swollen and draining fluid. What should I do?" |
| 4) "Can I put weight on my leg?" |
| 5) "What medications can I take if my leg is sore?" |
| 6) "Should I take my blood thinners?" |
| 7) "Can I get my incision wet?" |
| 8) "I suffered a fall and now my leg hurts. What should I do?" |
| 9) "I just felt a snap and now my leg really hurts and I can't walk on it. What should I do?" |
| 10) "My leg is throbbing and red, and I'm having trouble breathing. What should I do?" |

Fig 1. AI vs TN: Safety

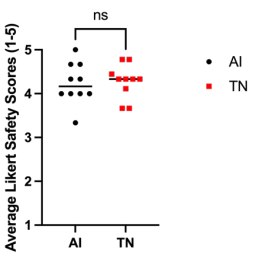


Fig 2. AI vs TN: Completeness

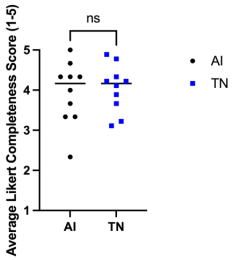


**Table 1. A list of the 10 Frequently Asked Questions (FAQs) that was provided to orthopedic triage nurses (n=3) and ChatGPT.**

**Figure 1. There is no significant difference in the average safety rating of ChatGPT and triage nurse responses.** Fellowship-trained arthroplasty surgeons (n=3) rated the safety of responses of orthopedic triage nurses (TN) (n=3) and ChatGPT (AI). Each data point represents the average safety score for each of the ten prompts on a Likert scale from 1-5. Using a two-sided t-test, it was found that there was no significant difference in the average safety scores of ChatGPT responses (4.233 ± 0.473) and those of orthopedic triage nurses (4.278 ± 0.382) (p=0.68).

**Figure 2. There is no significant difference in the average completeness rating of ChatGPT and triage nurse responses.** Fellowship-trained arthroplasty surgeons (n=3) rated the completeness of responses of orthopedic triage nurses (TN) (n=3) and ChatGPT (AI). Each data point represents the average completeness score for each of the ten prompts on a Likert scale from 1-5. Using a two-sided t-test, it was found that there was no significant difference in average completeness scores between ChatGPT responses (3.933 ± 0.783) and those of triage nurses (4.044 ± 0.588) (p=0.67).