

Evaluating the Quality and Usability of Artificial Intelligence–Generated Responses to Common Patient Questions in Foot and Ankle Surgery

Foot & Ankle Orthopaedics
2023, Vol. 8(4) 1–7
© The Author(s) 2023
DOI: 10.1177/24730114231209919
journals.sagepub.com/home/fao

Albert Thomas Anastasio, MD¹ , Frederic Baker Mills IV, MD, MS¹,
Mark P. Karavan, Jr, BS¹ , and Samuel B. Adams, Jr, MD¹

Abstract

Background: Artificial intelligence (AI) platforms, such as ChatGPT, have become increasingly popular outlets for the consumption and distribution of health care–related advice. Because of a lack of regulation and oversight, the reliability of health care–related responses has become a topic of controversy in the medical community. To date, no study has explored the quality of AI-derived information as it relates to common foot and ankle pathologies. This study aims to assess the quality and educational benefit of ChatGPT responses to common foot and ankle–related questions.

Methods: ChatGPT was asked a series of 5 questions, including “What is the optimal treatment for ankle arthritis?” “How should I decide on ankle arthroplasty versus ankle arthrodesis?” “Do I need surgery for Jones fracture?” “How can I prevent Charcot arthropathy?” and “Do I need to see a doctor for my ankle sprain?” Five responses (1 per each question) were included after applying the exclusion criteria. The content was graded using DISCERN (a well-validated informational analysis tool) and AIRM (a self-designed tool for exercise evaluation).

Results: Health care professionals graded the ChatGPT-generated responses as bottom tier 4.5% of the time, middle tier 27.3% of the time, and top tier 68.2% of the time.

Conclusion: Although ChatGPT and other related AI platforms have become a popular means for medical information distribution, the educational value of the AI-generated responses related to foot and ankle pathologies was variable. With 4.5% of responses receiving a bottom-tier rating, 27.3% of responses receiving a middle-tier rating, and 68.2% of responses receiving a top-tier rating, health care professionals should be aware of the high viewership of variable-quality content easily accessible on ChatGPT.

Level of Evidence: Level III, cross sectional study.

Keywords: ankle arthritis, Charcot arthropathy, ankle arthroplasty, Jones fracture, ankle sprain, education, artificial intelligence, ChatGPT

Introduction

The use of artificial intelligence (AI) methodologies for various applications in orthopaedic foot and ankle surgery is growing.^{3,7-10,24} ChatGPT (Generative Pretrained Transformer) (Open AI, San Francisco, CA) was launched by OpenAI on November 30, 2022, and is a generative language model tool with the capacity to enable the public to converse with an AI network utilizing a text-based

platform. Early in 2023, ChatGPT was the fastest-growing consumer application to date, reaching

¹Department of Orthopaedic Surgery, Duke University Medical Center, Durham, NC, USA

Corresponding Author:

Mark P. Karavan Jr, BS, Department of Orthopaedic Surgery, Duke University Medical Center, 200 Trent Drive, Durham, NC 27710, USA.
Email: mark.karavan@duke.edu



more than 100 million users.⁴ ChatGPT is capable of providing answers to a vast array of questions, generating potential applications across multiple service-based fields. Already, there is considerable excitement regarding its use in medicine.^{13,16,17} Generative language tools could be the answer to the challenge of excessive medical documentation that has been attributed as a major cause of physician burn out.^{15,17} Modern AI systems have evolved to the extent where they exhibit remarkable capacity to replicate human abilities: ChatGPT scored higher than the current passing threshold for third-year medical students on the National Board of Medical Examiners (NBME) Step 1 sample examination.⁶ Despite the recency of the launch of ChatGPT, its use within the orthopaedic community has already been proposed.¹⁶

Moreover, the impressive verbal capacities of ChatGPT and other generative language AIs may lead to a heightened trust in the responses generated from these applications than those that can be found through a simple internet search. Therefore, if the quality of the medical information contained within a ChatGPT-generated answer is poor, use of the application may ultimately lead to wide distribution of medical misinformation. Left to their own devices, patients will often seek sources that are inaccurate, leaving them misinformed.

Thus, the purpose of this study was to evaluate the quality and usability of answers generated using ChatGPT to a variety of questions pertaining to foot and ankle orthopaedic surgery. The generated answers were scored by fellowship-trained orthopaedic foot and ankle surgeons and by current accredited orthopaedic foot and ankle fellows.

Methods

Search Strategy and Data Collection

The artificial intelligence platform ChatGPT (<https://chat.openai.com/chat>) was asked to answer 5 commonly asked questions related to the foot and ankle on March 15, 2023. The search was conducted by using the following questions: “What is the optimal treatment for ankle arthritis?” “How should I decide on ankle arthroplasty versus ankle arthrodesis?” “Do I need surgery for Jones fracture?” “How can I prevent Charcot arthropathy?” and “Do I need to see a doctor for my ankle sprain?” Our intent was to analyze the responses that ChatGPT would give if patients were to ask the platform for advice related to these disorders of the foot and ankle. The search yielded 5 answers total, 1 per question (n=5).

Scoring System

Two separate scoring systems were used to evaluate the quality and educational value of the ChatGPT-generated

responses: the DISCERN, a previously validated tool used to determine reliability and quality of a treatment approach, and the AI Response Metric (AIRM), a metric devised by our research team to assess the quality of an AI-generated response to a medical question. The AIRM score was modified from a similar scale created for evaluating the utility of the social media platform TikTok as an educational tool for pathologies of the foot and ankle from previously published work by Tabarestani et al.¹⁹

DISCERN for reliability and quality assessment

The DISCERN is a questionnaire metric for written medical information that aims to assess the accuracy and relevancy of textual descriptions of health care treatment-related information. The tool is well validated, has been used since the late 1990s, and is composed of 16 questions (Table 1).

The first 8 questions listed in the DISCERN system assess reliability of the publication (DISCERN 1) and the next 7 questions determine the quality of the author’s source base (DISCERN 2). The final question then rates the publication as a whole in terms of its quality as a source of information (DISCERN 3). Although its original intended purpose was designed to be a tool for written information, it has been successfully adapted as a tool for scoring the quality of videos.¹² DISCERN scores are categorized as the following: excellent is denoted by scores of 63 to 75 points, good is denoted by scores of 51 to 62 points, fair is denoted by scores of 39 to 50 points, poor is denoted by scores of 27 to 38 points, and very poor is denoted by scores of 16 to 26 points.

AIRM for Educational Suitability Assessment

To grade the educational value of the ChatGPT-generated responses, we developed the AIRM as a modification from previously published work.¹¹ This test determines whether or not the viewers can properly understand and rely on the advice following reading ChatGPT’s response. The AIRM has 4 focuses, with a scale of 1 to 5 as grading options. Surgeons were tasked with selecting which set of answers most closely aligned with their thoughts about each answer provided by ChatGPT. Responses were graded using the metric in Table 2. Higher scores denote higher overall quality of ChatGPT-generated responses.

Assessment

The responses were collected by 2 reviewers and independently evaluated by our orthopaedic research team. Once the data regarding the artificial intelligence response metrics was collected for each response, the content of the videos was graded using the DISCERN and AIRM tools by a multiinstitutional group of foot and

Table 1. DISCERN Instrument.

Based on the answers to all the following questions, rate the overall quality of the publication as a source of information about treatment choices.

1 = serious or extensive shortcomings

2

3 = potentially important but not serious shortcomings

4

5 = minimal shortcomings

Are the aims clear?

Does it achieve its aims?

Is it relevant?

Is it clear what sources of information were used to compile the publication (other than the author or producer)?

Is it clear when the information used or reported in the publication was produced?

Is it balanced and unbiased?

Does it provide details of additional sources of support and information?

Does it refer to areas of uncertainty?

Does it describe how each treatment works?

Does it describe the benefits of each treatment?

Does it describe the risks of each treatment?

Does it describe what would happen if no treatment is used?

Does it describe how the treatment choices affect overall quality of life?

Is it clear that there may be more than one possible treatment choice?

Does it provide support for shared decision-making?

ankle orthopaedic surgeons and fellows. Each response was graded separately by 2 trained reviewers. Any points of discrepancy between the 2 reviewers were resolved by a third author.

Statistical Analysis

Scoring and characteristic data are presented as the mean (SD), median (interquartile range [IQR]), and percentage. A 2-sample *t* test was used to compare the 2 types of uploaders by using the mean, SD, and sample size of each continuous and categorical variable. Statistical significance was set at $P < .05$, for comparisons other than interrater reliability. All analyses were performed using Microsoft Excel (Microsoft Corp, Redmond, WA).

Results

Basic Characteristics

In total, 5 responses were given after using the aforementioned search questions into ChatGPT. Of the 5 responses given, all were included in the final analysis. Table 1

Table 2. AIRM Instrument.

“This response is something I would not be comfortable with my patient reading. This response is clearly not in line with the current literature consensus regarding this topic. This response is not clear with regards to grammar and syntax. This response is incomplete, and does not cover the topic in an appropriately nuanced manner.”

“This response is something I would not be comfortable with my patient reading.

This response is partially in line with the current literature consensus regarding this topic.

This response is not clear with regards to grammar and syntax. This response is incomplete, and covers the topic in a basic manner.”

“This response is potentially something I would be comfortable with my patient reading. This response is somewhat in line with the current literature consensus regarding this topic. This response is somewhat clear with regards to grammar and syntax. This response is partially complete, and covers the topic in a basic manner.”

“This response is something I would be comfortable with my patient reading.

This response is mostly in line with the current literature consensus regarding this topic.

This response is clear with regards to grammar and syntax.

This response is mostly complete, and covers the topic in a somewhat appropriately nuanced manner.”

“This response is something I would be comfortable with my patient reading.

This response is clearly in line with the current literature consensus regarding this topic.

This response is clear with regards to grammar and syntax. This response is complete, and covers the topic in an appropriately nuanced manner.”

presents the basic characteristics of the questions asked and the responses given by ChatGPT (Tables 3 and 4, Supplementary Tables 1–3).

Rating of Responses, DISCERN Questionnaire

Fellowship-trained orthopaedic foot and ankle surgeons and foot and ankle fellows were asked to grade the responses generated by ChatGPT to each question. Scoring for each question using the DISCERN questionnaire is described in Supplementary Table 4.

Rating of Responses, AIRM

Fellowship-trained orthopaedic foot and ankle surgeons and foot and ankle fellows were asked to grade the responses generated by ChatGPT to each question. Scoring for each question using the DISCERN questionnaire is described in Supplementary Table 5.

Table 3. Question and Answer 1.

Question 1:

“What is the optimal treatment for ankle arthritis?”

Chat GPT answer 1:

The optimal treatment for ankle arthritis depends on several factors, including the severity of the arthritis, the patient's age and activity level, and the location of the arthritis within the ankle joint. Here are some general treatments that may be recommended:

Non-surgical treatments: For mild to moderate ankle arthritis, non-surgical treatments may be recommended. These can include physical therapy, activity modification, orthotics, anti-inflammatory medications, and corticosteroid injections.

Ankle arthroscopy: For some patients with ankle arthritis, arthroscopic surgery may be recommended. This minimally invasive procedure uses a small camera and specialized tools to remove damaged tissue or bone spurs, smooth the joint surface, and improve joint function.

Ankle fusion (arthrodesis): In some cases of advanced ankle arthritis, ankle fusion may be recommended. This procedure involves fusing the ankle joint bones together, which eliminates the joint and relieves pain.

Ankle replacement (arthroplasty): In select patients with advanced ankle arthritis, ankle replacement may be recommended. This involves replacing the damaged joint with an artificial joint.

The best treatment for ankle arthritis will depend on your individual circumstances. An orthopedic surgeon can evaluate your condition and recommend the most appropriate treatment plan.

Table 4. Question and Answer 2.

Question 2:

“How should I decide on ankle arthroplasty versus ankle arthrodesis?”

ChatGPT Answer 2

“The decision between ankle arthroplasty (replacement) and ankle arthrodesis (fusion) depends on several factors, including the patient's age, activity level, overall health, and the severity and location of the ankle joint damage. Here are some general considerations for both procedures:

Ankle Arthroplasty:

This procedure involves replacing the damaged joint with an artificial joint, usually made of metal and plastic components. It is typically recommended for patients who have relatively normal bone alignment and minimal arthritis in the surrounding joints.

Ankle arthroplasty may provide greater mobility and flexibility in the ankle joint compared to ankle arthrodesis, which fuses the joint.

Ankle Arthrodesis:

This procedure involves fusing the ankle joint to prevent movement and pain.

It is typically recommended for patients who have severe joint damage, deformity, or instability, or those who have failed previous ankle surgeries.

Ankle arthrodesis typically results in a stable, pain-free joint, but limits ankle mobility.

Ultimately, the decision to have ankle arthroplasty versus ankle arthrodesis should be made in consultation with an orthopedic surgeon, who can evaluate your individual circumstances and recommend the best course of treatment.”

Discussion

This study was the first in orthopaedic foot and ankle surgery to assess the quality and usability of answers generated using ChatGPT to a variety of common patient questions. We found a high degree of interobserver variation in the assessment of the answers generated by ChatGPT. Several evaluators were impressed with the ChatGPT responses, scoring them very highly. Others found the answers to be deficient in a number of categories and scored the responses at lower- to middle-tier ratings. This preliminary analysis is underpowered and lacks the validation required to recommend the widespread use of ChatGPT in clinical practice. Rather, our intention was to generate discussion regarding the use cases of AI systems in orthopaedic foot and ankle surgery. Several of the foot and ankle fellows and attending surgeons at our institution were unaware of ChatGPT and remarked at the impressive grammatical correctness of the reviewed answers as well as the perceived completeness of the response.

This study is the first to evaluate ChatGPT-generated responses to several common questions related to foot and ankle orthopaedics. Other studies have investigated the usage of this AI modality in other areas of medicine. A recent systematic review identified 65 and 100 papers in the PubMed and EuropePMC databases, respectively. An additional systematic review of the use of ChatGPT in health care identified 60 manuscripts for inclusion. Advantages to the use of ChatGPT were listed in 51 of the 60 studies (85.0%). Benefits of ChatGPT included efficient analysis of data sets and performance of literature reviews within research, workflow streamlining and cost saving in the clinical setting, and improvements in health care education including improved personalized learning.¹⁸

Although ChatGPT exhibited numerous avenues for potential benefit within health care, 58 of the 60 included studies (96.7%) reported concerns regarding use of this technology. Copyright and plagiarism ethics, citation and reference transparency, and inaccurate content were all cited as potential drawbacks to the use of ChatGPT in health care.¹⁸ Several attending surgeons at our institution commented on failure of ChatGPT to include in-text citations, leading to a lack of granularity regarding source material.

Our study is the first to evaluate ChatGPT in foot and ankle. However, other AI models have been previously researched within domains relevant to foot and ankle surgery. Radiographic interpretation is an area of active research within computational modeling. Kitamura et al²⁵ utilized 5 different convolutional neural networks (CNNs) to detect ankle fractures from plain radiographs. They internally validated each of the CNNs and noted a fair CNN fracture detection accuracy of 81%. Ashkani-Esfahani et al² internally validated 2 deep CNNs for the same purpose and achieved a near-perfect area under the curve of 0.99. AI has been used for image analysis throughout foot and ankle

surgery, including for Achilles tendinopathy,²¹ stress fracture,²² Lisfranc fracture,¹ and calcaneal fracture.⁵

Symptom, complication, and outcome prediction is another area where AI is applied broadly in foot and ankle surgery. Support vector machine modeling has been applied for classifying hallux valgus patients as having painful feet or pain-free feet. Using radiographic metrics such as hallux valgus angle (HVA), intermetatarsal angle (IMA), and distal metatarsal articular angle (DMAA), Wang et al²⁰ internally validated this approach to hallux valgus classification, with an accuracy of 76.4%. Likewise, logistic regression testing and gradient boosting models have been used to predict short-term complications such as readmissions and mortality following open reduction and internal fixation for ankle fractures. Merrill et al¹⁴ internally validated these approaches and found both models to perform similarly, with areas under the curves for logistic regression ranging from 0.7101 to 0.7583 and areas under the curves for gradient boosting ranging from 0.6979 to 0.7580. A recent systematic review by Gupta et al⁷ provides an overview of 31 studies related to the use of AI in foot and ankle surgery and is an excellent resource for surgeons interested in learning about the capabilities of AI in our field.

A considerable potential exists for further investigation of generative-text AI systems such as ChatGPT within foot and ankle surgery. Direct comparison of the ChatGPT-generated responses to answers curated from simple internet search tools would elucidate whether newer-model AI provides a valuable addition to what is currently widely available on search platforms. In addition, research should seek to characterize the accuracy, reliability, and usability of ChatGPT-generated clinical documentation. Is this technology adequate for operative report generation and coding and billing purposes? Could ChatGPT combined with voice recognition software replace in-person or virtual scribes in a high-volume clinical setting? The answer to these questions has the potential to revolutionize the practice of surgery and medicine in an era of increased burnout,²³ freeing up provider time for direct patient care.

This study has several important limitations. Most notably, our study samples respondents from a single institution, including fellows and attending foot and ankle surgeons who may have particular, institutional practice patterns that have influenced the evaluation of the responses generated by ChatGPT. As an additional limitation, we used a nonvalidated tool of our own creation to assess the quality of responses generated by ChatGPT. Further research should seek to validate metrics that can be used to assess the usability of AI-generated text blocks. We sought to minimize these limitations by using a well-validated and heavily applied tool for evaluation of written medical text, the DISCERN. Finally, this study was completed within several weeks after the initial release of ChatGPT. Subsequent models may be further refined as the AI algorithm continues to develop and learn.

Conclusion

In summary, ChatGPT is a rapidly growing generative language tool with the capacity to provide answers to a wide variety of common patient questions. This study was the first in orthopaedic foot and ankle surgery to assess the quality and usability of answers generated using ChatGPT to a variety of common patient questions in our field. We found a high degree of interobserver variation in the assessment of the answers generated by ChatGPT, with several evaluators rating the responses very favorably, and others critiquing the ChatGPT-generated answers. Generative language AI systems have the potential to greatly impact the provision of medical care, and further investigation will be required to assess the various use cases of this technology across foot and ankle surgery.

Ethical Approval

No ethical approval was needed for this study. No patients were included in this work. Online, publicly available Artificial Intelligence responses were analyzed.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. ICMJE forms for all authors are available online.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Albert Thomas Anastasio, MD,  <https://orcid.org/0000-0001-5817-3826>

Mark P. Karavan Jr, BS,  <https://orcid.org/0000-0002-8954-5861>

References

1. Ashkani-Esfahani S, Mojahed-Yazdi R, Bhimani R, et al. Deep learning algorithms improve the detection of subtle Lisfranc malalignments on weightbearing radiographs. *Foot Ankle Int.* 2022;43(8):1118-1126.
2. Ashkani-Esfahani S, Yazdi RM, Bhimani R, et al. Detection of ankle fractures using deep learning algorithms. *Foot Ankle Surg.* 2022;28(8):1259-1265.
3. Chande RD, Hargraves RH, Ortiz-Robinson N, Wayne JS. Predictive behavior of a computational foot/ankle model through artificial neural networks. *Comput Math Methods Med.* 2017;2017:3602928. doi:10.1155/2017/3602928
4. Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. *JMIR Med Educ.* 2023;9:e46885. doi:10.2196/46885
5. Farda NA, Lai JY, Wang JC, Lee PY, Liu JW, Hsieh IH. Sanders classification of calcaneal fractures in CT images

- with deep learning and differential data augmentation techniques. *Injury*. 2021;52(3):616-624.
6. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9:e45312. doi:10.2196/45312
 7. Gupta P, Kingston KA, O'Malley M, Williams RJ, Ramkumar PN. Advancements in artificial intelligence for foot and ankle surgery: a systematic review. *Foot Ankle Orthop*. 2023;8(1):24730114221151079. doi:10.1177/24730114221151079
 8. Hernigou P, Olejnik R, Safar A, Martinov S, Hernigou J, Ferre B. Digital twins, artificial intelligence, and machine learning technology to identify a real personalized motion axis of the tibiotalar joint for robotics in total ankle arthroplasty. *Int Orthop*. 2021;45(9):2209-2217. doi:10.1007/s00264-021-05175-2
 9. Hernigou P, Scarlat MM. Ankle and foot surgery: from arthrodesis to arthroplasty, three dimensional printing, sensors, artificial intelligence, machine learning technology, digital twins, and cell therapy. *Int Orthop*. 2021;45(9):2173-2176. doi:10.1007/s00264-021-05191-2
 10. Islam M, Hsiao-Weckslar ET. Detection of gait modes using an artificial neural network during walking with a powered ankle-foot orthosis. *J Biophys*. 2016;2016:7984157. doi:10.1155/2016/7984157
 11. Jang CW, Kim M, Kang SW, Cho HE. Reliability, quality, and educational suitability of TikTok videos as a source of information about scoliosis exercises: a cross-sectional study. *Healthcare (Basel)*. 2022;10(9):1622. doi:10.3390/healthcare10091622
 12. Kuru T, Erken HY. Evaluation of the quality and reliability of YouTube videos on rotator cuff tears. *Cureus*. 2020;12(2):e6852. doi:10.7759/cureus.6852
 13. Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: current applications, future possibilities and limitations of ChatGPT. *Diagn Interv Imaging*. 2023;104(6):269-274. doi:10.1016/j.diii.2023.02.003
 14. Merrill RK, Ferrandino RM, Hoffman R, Shaffer GW, Ndu A. Machine learning accurately predicts short-term outcomes following open reduction and internal fixation of ankle fractures. *J Foot Ankle Surg*. 2019;58(3):410-416.
 15. Muhiyaddin R, Elfadl A, Mohamed E, et al. Electronic health records and physician burnout: a scoping review. *Stud Health Technol Inform*. 2022;289:481-484. doi:10.3233/shti210962
 16. Ollivier M, Pareek A, Dahmen J, et al. A deeper dive into ChatGPT: history, use and future perspectives for orthopaedic research. *Knee Surg Sports Traumatol Arthrosc*. 2023;31(4):1190-1192. doi:10.1007/s00167-023-07372-5
 17. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health*. 2023;5(3):e107-e108. doi:10.1016/S2589-7500(23)00021-3
 18. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)*. 2023;11(6):887. doi:10.3390/healthcare11060887
 19. Tabarestani TQ, Anastasio AT, Duruewuru A, Taylor JR, Bethell MA, Adams SB. Analyzing the quality and educational value of Achilles tendinopathy-related videos on TikTok. *Foot Ankle Surg*. 2023;29(4):350-354. doi:10.1016/j.fas.2023.03.007
 20. Wang C, Li C, Zhang R, et al. Identification of radiographic characteristics associated with pain in hallux valgus patients: a preliminary machine learning study. *Front Public Health*. 2022;10:943026.
 21. Wang L, Wen D, Yin Y, et al. Musculoskeletal ultrasound image-based radiomics for the diagnosis of Achilles tendinopathy in skiers. *J Ultrasound Med*. 2023;42(2):363-371.
 22. Wang Y, Li Y, Lin G, et al. Lower-extremity fatigue fracture detection and grading based on deep learning models of radiographs. *Eur Radiol*. 2023;33(1):555-565.
 23. West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions. *J Intern Med*. 2018;283(6):516-529. doi:10.1111/joim.12752
 24. Zhu S, Niu Y, Wang J, Xu D, Li Y. Artificial intelligence technology combined with ultrasound-guided needle knife interventional treatment of PF: improvement of pain, fascia thickness, and ankle-foot function in patients. *Comput Math Methods Med*. 2022;2022:3021320. doi:10.1155/2022/3021320
 25. Kitamura G, Chung CY, Moore BE 2nd. Ankle fracture detection utilizing a convolutional neural network ensemble implemented with a small sample, De Novo training, and multi-view incorporation. *J Digit Imaging*. 2019;32(4):672-677. doi:10.1007/s10278-018-0167-7.

Supplementary Table 1. Question and Answer 3.

Question 3:

“Do I need surgery for Jones fracture?”

Chat GPT Answer 3:

“The decision to undergo surgery for a Jones fracture depends on several factors, including the severity of the fracture, the location of the fracture, and the patient's age, activity level, and overall health. Here are some general considerations:

A Jones fracture is a type of fracture that occurs in the fifth metatarsal bone of the foot, typically in the middle or proximal part of the bone.

In some cases, a Jones fracture can heal on its own with conservative treatment, such as immobilization in a cast or walking boot.

However, in other cases, surgery may be necessary to promote healing and prevent complications, such as delayed healing, nonunion (failure to heal), or chronic pain.

Surgery for Jones fracture typically involves placing a screw or other hardware across the fracture site to stabilize the bone and promote healing.

Your orthopedic surgeon will evaluate your individual circumstances, including the severity and location of the fracture, as well as your overall health and activity level, to determine whether surgery is necessary.

It's important to consult with an orthopedic surgeon if you suspect you have a Jones fracture. The earlier the fracture is diagnosed and treated, the better the chances for successful healing and recovery.”

Supplementary Table 2. Question and Answer 4.

Question 4:

“How can I prevent Charcot arthropathy?”

Chat GPT Answer 4:

Charcot arthropathy, also known as Charcot foot or neuropathic arthropathy, is a condition that affects people with nerve damage in their feet, most commonly those with diabetes. Here are some tips to help prevent Charcot arthropathy:

Control your blood sugar: Maintaining good blood sugar control is one of the most important steps in preventing complications related to diabetes, including Charcot arthropathy. Work with your health care team to develop a diabetes management plan that works for you.

Inspect your feet daily: Check your feet every day for any changes in color, temperature, or sensation, as well as any cuts, blisters, or other injuries. Early detection and treatment of foot problems can help prevent Charcot arthropathy.

Wear proper footwear: Wear shoes that fit well and provide adequate support and cushioning. Avoid high-heeled shoes or shoes with pointed toes, as these can put excess pressure on your feet and increase your risk of developing Charcot arthropathy.

Protect your feet: Use caution when walking barefoot or wearing open-toed shoes, and always wear socks and shoes when outside. If you have decreased sensation in your feet, consider wearing protective shoes or socks to prevent injury.

Regular check-ups: Schedule regular check-ups with your healthcare team, including foot exams to check for signs of neuropathy or Charcot arthropathy.

Seek medical attention promptly: If you notice any changes in your feet, such as swelling, redness, or warmth, or if you experience any pain or discomfort, seek medical attention promptly.

By following these preventive measures, you can help reduce your risk of developing Charcot arthropathy and other foot complications related to diabetes.

Supplementary Table 3. Question and Answer 5.

Question 5:

“Do I need to see a doctor for my ankle sprain?”

ChatGPT Answer 5:

“If you have an ankle sprain, it is always a good idea to see a doctor, especially if it is severe or if you are experiencing significant pain or swelling. A doctor can assess the extent of the injury and determine the appropriate treatment plan to help you recover as quickly and safely as possible.

If you have a mild sprain, you can try home remedies such as rest, ice, compression, and elevation (RICE) to alleviate the symptoms.

However, if the pain and swelling persist or if you have trouble bearing weight on the affected ankle, you should seek medical attention. Additionally, if you experience any numbness, tingling, or loss of sensation in your foot or ankle, you should seek medical attention immediately, as these symptoms may indicate a more serious injury.”

Supplementary Table 4. DISCERN Questionnaire Grading of ChatGPT-Generated Responses to Questions 1-5 by Orthopaedic Foot and Ankle Surgeons and Fellows.

DISCERN Scores	Scorer 1	Scorer 2	Scorer 3	Scorer 4	Scorer 5	Scorer 6	Scorer 7	Scorer 8	Scorer 9	Scorer 10	Scorer 11
DISCERN score 1	60	37	60	38	38	60	60	15	75	45	75
DISCERN score 2	45	35	60	30	35	45	75	15	75	45	75
DISCERN score 3	45	42	60	30	41	60	75	15	75	45	75
DISCERN score 4	60	34	75	30	42	75	75	15	75	60	75
DISCERN score 5	45	31	60	30	32	75	75	30	75	60	75
DISCERN score average	51	35.8	63	31.6	37.6	63	72	18	75	51	75

Supplementary Table 5. AIRM Questionnaire Grading of ChatGPT-Generated Responses to Questions 1-5 by Orthopaedic Foot and Ankle Surgeons and Fellows.

AIRM Scores	Scorer 1	Scorer 2	Scorer 3	Scorer 4	Scorer 5	Scorer 6	Scorer 7	Scorer 8	Scorer 9	Scorer 10	Scorer 11
AIRM score 1	3	4	5	4	3	4	5	2	4	2	5
AIRM score 2	3	2	5	4	2	4	5	3	5	3	5
AIRM score 3	3	4	5	4	5	3	5	1	5	3	5
AIRM score 4	4	4	5	4	5	5	5	4	5	4	5
AIRM score 5	3	4	4	4	4	4	5	3	5	4	5
AIRM score total	16	18	24	20	19	20	25	13	24	16	25