

# Assessing the quality of ChatGPT's responses to commonly asked questions about trigger finger treatment

✉ Mehmet Can Gezer,<sup>1</sup> ✉ Mehmet Armangil<sup>2</sup>

<sup>1</sup>Department of Orthopedics and Traumatology, Mamak State Hospital, Ankara-Türkiye

<sup>2</sup>Department of Orthopedics and Traumatology, Hand Surgery Unit, Ankara University Faculty of Medicine, Ankara-Türkiye

## ABSTRACT

**BACKGROUND:** This study aims to evaluate the accuracy and reliability of Generative Pre-trained Transformer (ChatGPT; OpenAI, San Francisco, California) in answering patient-related questions about trigger finger. This evaluation has the potential to enhance patient education prior to treatment and provides insight into the role of artificial intelligence (AI)-based systems in the patient education process.

**METHODS:** The ten most frequently asked questions regarding trigger finger were compiled from patient education websites and a literature review, then posed to ChatGPT. Two orthopedic specialists evaluated the responses using the Journal of the American Medical Association (JAMA) Benchmark criteria and the DISCERN instrument (A Tool for Judging the Quality of Written Consumer Health Information on Treatment Choices). Additionally, the readability of the responses was assessed using the Flesch-Kincaid Grade Level.

**RESULTS:** The DISCERN scores for ChatGPT's responses to trigger finger questions ranged from 35 to 47, with an average of 42, indicating "moderate" quality. While 60% of the responses were satisfactory, 40% contained deficiencies. According to the JAMA Benchmark criteria, the absence of scientific references was a significant drawback. The average readability level corresponded to the university level, making the information difficult to understand for patients with low health literacy. Improvements are needed to enhance the accessibility and comprehensibility of the content for a broader patient population.

**CONCLUSION:** To the best of our knowledge, this is the first study to investigate the use of ChatGPT in the context of trigger finger. While ChatGPT shows reasonable effectiveness in providing general information on trigger finger, expert oversight is necessary before it can be relied upon as a primary source for patient education.

**Keywords:** Trigger finger; ChatGPT; DISCERN; patient education; artificial intelligence.

## INTRODUCTION

In recent years, rapid advancements in artificial intelligence (AI) technologies have introduced new opportunities in the healthcare sector, particularly in enhancing patient education and information dissemination. At the forefront of these innovations is the Generative Pre-trained Transformer (ChatGPT; OpenAI, San Francisco, CA), a large-scale language model capable of generating quick, contextually relevant, and personalized responses to medical inquiries. As patients increasingly turn to digital platforms to obtain information about

diagnoses and treatment options, AI-powered tools like ChatGPT offer significant advantages in terms of accessibility, immediacy, and scalability of medical information. However, the reliability, accuracy, and clinical applicability of these tools in patient education remain areas in need of thorough evaluation, especially when applied to specific medical conditions such as trigger finger. As such tools continue to gain prominence, it becomes essential to assess their potential role in supplementing traditional patient-physician communication and their overall value in modern medical practice.

Cite this article as: Gezer MC, Armangil M. Assessing the quality of ChatGPT's responses to commonly asked questions about trigger finger treatment. *Ulus Travma Acil Cerrahi Derg* 2025;31:389-393.

Address for correspondence: Mehmet Can Gezer

Department of Orthopedics and Traumatology, Mamak State Hospital, Ankara, Türkiye

E-mail: mehmetcangezer9121@gmail.com

*Ulus Travma Acil Cerrahi Derg* 2025;31(4):389-393 DOI: 10.14744/tjtes.2025.32735

Submitted: 08.01.2025 Revised: 17.01.2025 Accepted: 31.01.2025 Published: 28.03.2025

OPEN ACCESS This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).



Trigger finger (stenosing tenosynovitis) is a common condition characterized by the thickening and swelling of the flexor tendons within their synovial sheath, resulting in restricted finger movement. This condition often causes a sensation of catching or locking during flexion, limiting the smooth motion of the affected finger.<sup>[1]</sup> Trigger finger is more commonly observed in individuals who engage in repetitive hand movements or those with comorbid conditions such as diabetes.<sup>[2,3]</sup> The symptoms include finger catching, pain, and stiffness, all of which negatively impact the patient's ability to perform daily activities.<sup>[4]</sup> Therefore, it is critically important for patients to have access to accurate and comprehensible information regarding the treatment of trigger finger.

Evaluating how accurately and satisfactorily AI-based systems, such as ChatGPT, can respond to patient inquiries regarding trigger finger is important for healthcare services. This study aims to assess the accuracy and reliability of ChatGPT's responses to patient questions about trigger finger. The use of AI-based tools in the patient education process may contribute to more active patient involvement in their treatment and enable them to make more informed decisions.

## MATERIALS AND METHODS

This study aims to evaluate the responses provided by ChatGPT to patient inquiries regarding trigger finger. The study is designed to assess the scientific validity and appropriateness of the responses in the context of patient education. The data used in this study were collected from the ten most frequently asked questions, compiled from patient education websites and scientific literature. These questions cover a range of topics, including the definition of trigger finger, its symptoms, treatment options, surgical considerations, risk factors, and rehabilitation processes.

The questions were posed to ChatGPT in a single session without providing any additional information. No human intervention or supplementary clarification was offered during the generation of responses. The responses were recorded and subsequently evaluated by two orthopedic specialists. The specialists assessed the accuracy, comprehensiveness, clarity, and evidence-based nature of the responses using the Journal of the American Medical Association Benchmark (JAMA) Benchmark criteria and the DISCERN instrument (A Tool for Judging the Quality of Written Consumer Health Information on Treatment Choices). DISCERN is a widely used tool for evaluating the quality of health information, particularly in determining whether the information is unbiased and balanced. The DISCERN instrument consists of 15 questions,<sup>[5]</sup> each scored on a scale from 1 to 5, resulting in a cumulative score ranging from 15 to 75. The tool is divided into two sections: the first section (questions 1-8) evaluates the reliability of the publication, while the second section (questions 9-15) assesses the relevance of the treatment information provided. The DISCERN score classification system is as

follows: scores between 63 and 75 are rated as "excellent," 51 to 62 as "good," 39 to 50 as "average," 28 to 38 as "poor," and scores below 28 are considered "very poor." Higher scores indicate higher quality of information.

The quality of the information was evaluated using the DISCERN score and the JAMA Benchmark criteria, in line with previous studies.<sup>[5,6]</sup> The JAMA Benchmark criteria assess the quality of health-related information based on four key components: authorship, attribution, disclosure, and currency.<sup>[7]</sup> Each criterion was clearly defined, and adherence to these standards was ensured during the evaluation. Evaluators assigned a score of "1" for each item that fulfilled the respective criterion.<sup>[7,8]</sup>

The readability of the responses was analyzed using the Flesch-Kincaid Grade Level test. This test assesses the complexity of written text by considering the number of syllables per word and the number of words per sentence. A higher score indicates that the text is easier to read. According to the scoring system, a score between 0 and 6 represents a basic reading level, a score between 6 and 12 indicates an average reading level, and scores between 12 and 18 suggest an advanced reading level. Grade levels, such as "6th Grade" or "12th Grade" are used to indicate the educational level required to understand the text.<sup>[9]</sup>

The scores from the DISCERN, JAMA Benchmark, and Flesch-Kincaid Grade Level tests were averaged based on the evaluations of two reviewers. No discrepancies were observed between the evaluators.

### Statistical Analysis

The data were analyzed using the Statistical Package for the Social Sciences (SPSS) version 20. To assess the consistency of the ratings provided by the two orthopedic specialists, Cohen's Kappa coefficient was calculated. Cohen's Kappa is a statistical method used to measure inter-rater agreement, and in this study, it was employed to evaluate the reliability of the assessment process. The analysis using Cohen's Kappa coefficient showed a 95% agreement between the evaluators.

## RESULTS

The DISCERN scores for ChatGPT's responses to frequently asked questions about trigger finger ranged from 35 to 47, with an average score of 42. This indicates that the overall quality of information was "moderate." While 60% of the responses were considered moderately satisfactory, 40% exhibited deficiencies and required further clarification. The evaluation using the DISCERN tool showed that most responses were adequate for providing general information; however, they lacked sufficient detail and evidence-based references (Table 1).

According to the JAMA Benchmark criteria, none of the responses included direct evidence-based references. The av-

**Table 1.** Evaluation of ChatGPT's responses

Question	Response Quality	DISCERN Score	Flesch-Kincaid Grade Level
What is trigger finger, and what are its symptoms?	Moderate	43	13.5
What are the treatment options for trigger finger?	Moderate	45	14.0
When is surgery for trigger finger necessary?	Poor	39	14.2
What should the rehabilitation process for trigger finger be like?	Moderate	47	13.3
What are the risks of trigger finger surgery?	Poor	41	14.1
What are the causes of trigger finger?	Moderate	42	13.9
Can trigger finger be prevented?	Poor	38	13.8
How effective is steroid injection for trigger finger?	Moderate	42	14.3
Is rehabilitation necessary after trigger finger surgery?	Moderate	44	13.6
What is the risk of complications after trigger finger surgery?	Poor	35	14.0

erage JAMA Benchmark score was 1.2. The results of the Flesch-Kincaid Grade Level test indicated that the average readability level of ChatGPT's responses was 14.0.

### Evaluation of Responses

**Question:** What is trigger finger, and what are its symptoms?

**Answer:** Trigger finger is a condition characterized by the thickening of the flexor tendons in the hand and the narrowing of the synovial sheath, resulting in catching or locking sensations during finger movements. Clinical signs include finger stiffness, which may be more pronounced in the morning, pain in the finger joints, and a sensation of "catching" during flexion. In the early stages, symptoms may be intermittent, but as the condition progresses, permanent finger flexion or restricted movement may develop.

**DISCERN Score:** 43

**Response Quality:** Moderate

**Question:** What are the treatment options for trigger finger?

**Answer:** The initial treatment for trigger finger typically involves conservative methods. These include resting the hand, using splints to restrict movement, and the use of nonsteroidal anti-inflammatory drugs (NSAIDs). Corticosteroid injections are also commonly used to reduce inflammation and manage symptoms. Surgical treatment is considered when conservative methods fail and involves releasing the tendon sheath to allow the finger to move freely.

**DISCERN Score:** 45

**Response Quality:** Moderate

**Question:** When is surgery for trigger finger necessary?

**Answer:** Surgery for trigger finger is required when conservative treatment methods (e.g., splint use, NSAIDs, and corticosteroid injections) fail to adequately control symptoms. Surgical intervention is preferred in cases of severe pain and restricted movement that significantly affect the patient's daily activities. The goal of surgery is to enlarge the synovial

sheath to allow the tendon to move freely.

**DISCERN Score:** 39

**Response Quality:** Poor

**Question:** What should the rehabilitation process for trigger finger be like?

**Answer:** The rehabilitation process following trigger finger surgery aims to restore hand and finger function. In the early stages, gentle motion exercises are recommended to support soft tissue healing in the hand. Physical therapy plays a crucial role in strengthening the hand muscles, increasing joint range of motion, and reducing swelling. Adherence to post-operative recommendations during the rehabilitation process will accelerate the recovery period.

**DISCERN Score:** 47

**Response Quality:** Moderate

**Question:** What are the risks of trigger finger surgery?

**Answer:** Postoperative complications following trigger finger surgery may include infection, bleeding, nerve damage, delayed wound healing, and, in rare cases, permanent finger stiffness. Additionally, damage to the digital nerve during surgery can result in sensory loss in the finger. These risks may vary depending on the surgical technique, the surgeon's experience, and the patient's overall health condition.

**DISCERN Score:** 41

**Response Quality:** Poor

**Question:** What are the causes of trigger finger?

**Answer:** The most common causes of trigger finger include overuse of the fingers, repetitive hand movements, prolonged hand strain, and inflammatory conditions such as diabetes and rheumatoid arthritis. These factors lead to thickening and inflammation of the tendon sheath, making it difficult for the tendon to move freely. Additionally, trigger finger is more commonly observed in women over 40 years of age and in professions that require intensive use of the hands.

**DISCERN Score: 42****Response Quality:** Moderate**Question:** Can trigger finger be prevented?

**Answer:** Although trigger finger cannot be completely prevented, the risk can be significantly reduced by minimizing overuse of the hand and avoiding repetitive strenuous movements. In particular, maintaining ergonomic hand positions and preventing overuse of the fingers is important. Strengthening hand muscles and performing regular stretching exercises can also help reduce the risk.

**DISCERN Score: 38****Response Quality:** Poor**Question:** How effective is steroid injection for trigger finger?

**Answer:** Steroid injection is a commonly used method in the treatment of trigger finger to reduce inflammation and manage symptoms. Steroid injections can be highly effective, particularly in early-stage trigger finger cases, and can lead to complete resolution of symptoms in a large proportion of patients. However, it has been noted that the effectiveness may decrease with repeated injections, and the risk of complications may increase in the long term.

**DISCERN Score: 42****Response Quality:** Moderate**Question:** What is the recovery process after trigger finger surgery?

**Answer:** The recovery process after trigger finger surgery is generally quick, and most patients can return to normal activities within a few weeks. In the first few days, it is important to keep the hand elevated and apply ice to control swelling and pain. Stitches are typically removed 10-14 days after surgery. Physical therapy and home exercises help restore the normal range of motion and strength of the hand. The total recovery time may vary depending on the patient's overall health and the extent of the surgery.

**DISCERN Score: 44****Response Quality:** Moderate**Question:** What is the risk of complications after trigger finger surgery?

**Answer:** The risk of complications following trigger finger surgery is low, but it cannot be entirely eliminated. Potential complications include infection, nerve damage, delayed wound healing, and, rarely, permanent finger stiffness. Proper postoperative care and hygiene measures can minimize these risks. In rare cases, damage to the digital nerve may result in sensory loss in the finger. The risk of complications can vary depending on the surgeon's experience and the patient's adherence to postoperative recommendations.

**DISCERN Score: 35****Response Quality:** Poor**DISCUSSION**

ChatGPT demonstrates a certain level of success in providing general information about trigger finger. The moderate DISCERN scores indicate that the overall quality of the information is acceptable, though not without limitations. However, the absence of scientific references represents a significant gap in terms of reliability. According to the evaluation based on the JAMA Benchmark criteria, ChatGPT's responses lack a scientific foundation. This limitation is a critical concern in the context of patient education.

In this study, the average benchmark score was found to be 1.2, indicating a significant deficiency in the quality and reliability of the information provided for patient education. Although the responses effectively delivered general medical information, the lack of support from scientific sources diminishes their reliability and may hinder patients' ability to access reliable information. This limitation becomes even more critical in situations involving complex treatment options or surgical interventions.

The readability level of the responses corresponds to a university-level reading difficulty, making it challenging for the general patient population to understand the information. Individuals with low health literacy may struggle to comprehend and apply this content effectively. Therefore, it is important to present the responses in simpler and more comprehensible language.

The quality and accuracy of ChatGPT's responses are not considered sufficient for patient education. However, it is believed that AI-based tools like this can still be beneficial by providing patients with quick access to basic information and helping them develop a foundational understanding before consultations with their physicians. It is clear that the information provided by ChatGPT should be verified and, when necessary, expanded by specialist physicians.

Several recommendations can be made to enhance ChatGPT's potential in the patient education process. Firstly, presenting responses in simpler language and adjusting them to a lower reading level would improve their comprehensibility for the general patient population. Additionally, supporting responses with scientific references and citing these sources would enhance the reliability of the information. Before AI tools like ChatGPT are used as primary sources for patient education, it is crucial that healthcare professionals review the accuracy and reliability of the content. Such oversight can help ensure patients receive accurate information, improve health literacy, and support more informed participation in the treatment process.

In conclusion, ChatGPT has potential in providing general information about trigger finger. However, to fully realize this potential in patient education, the scientific basis of its responses needs to be strengthened, and the reading level should be lowered. AI-based tools like this can play a supportive role in the patient education process; however, the accuracy and reliability of the information provided by such tools must be validated by healthcare professionals.

## Limitations

This study has several important limitations. First, the number of questions used in the study was limited and does not cover all possible inquiries that patients may have about trigger finger. An evaluation based on only ten questions may not fully reflect ChatGPT's overall performance. Second, only the opinions of two orthopedic specialists were considered in the evaluation of the responses, which may have introduced subjective differences that could influence the results. Third, the timeliness and validity of ChatGPT's responses may change over time, meaning that similar results may not be replicated in future assessments. Finally, this study focused solely on the linguistic and content aspects of ChatGPT's responses and did not evaluate their actual impact on patient behavior. Future research should more comprehensively explore the effects of AI-based tools on patient education and engagement in the treatment process.

## CONCLUSION

This study demonstrates that ChatGPT's responses to patient questions regarding trigger finger are generally at an acceptable level but are limited by the lack of scientific references and a high reading difficulty. Therefore, it is concluded that the information provided by ChatGPT should be supplemented by expert input before being used as a primary source in the patient education process.

**Ethics Committee Approval:** Since this study is based on artificial intelligence, it does not require ethical committee approval, unlike traditional research involving human participants.

**Peer-review:** Externally peer-reviewed.

**Authorship Contributions:** Concept: M.C.G.; Design: M.C.G.; Supervision: M.A.; Resource: M.C.G.; Materials:

M.C.G.; Data Collection and/or Processing: M.A.; Analysis and/or Interpretation: M.A.; Literature Review: M.A.; Writing: M.C.G.; Critical Review: M.A.

**Conflict of Interest:** None declared.

**Financial Disclosure:** The author declared that this study has received no financial support.

## REFERENCES

- Jordaan PW, Klumpp R, Zeppieri M. Triggering, clicking, locking and crepitus of the finger: A comprehensive overview. *World J Orthop* 2023;14:733–40. [CrossRef]
- Pompeu Y, Aristega Almeida B, Kunze K, Altman E, Fufa DT. Current concepts in the management of advanced trigger finger: A critical analysis review. *JBJS Rev* 2021;9:e21. [CrossRef]
- Ferrara PE, Codazza S, Cerulli S, Maccauro G, Ferriero G, Ronconi G. Physical modalities for the conservative treatment of wrist and hand's tenosynovitis: A systematic review. *Semin Arthritis Rheum* 2020;50:1280–90. [CrossRef]
- Aspinen S, Nordback PH, Anttila T, Stjernberg-Salmela S, Ryhänen J, Kosola J. Platelet-rich plasma versus corticosteroid injection for treatment of trigger finger: Study protocol for a prospective randomized triple-blind placebo-controlled trial. *Trials* 2020;21:984. [CrossRef]
- Charnock D, Shepperd S, Needham G, Gann R. DISCERN: An instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* 1999;53:105–11. [CrossRef]
- Yoon SJ, Kim JB, Jung KJ, Chang HJ, Hong YC, Hong CH, et al. Evaluation of the quality of information available on the internet regarding chronic ankle instability. *Medicina* 2022;58:1315. [CrossRef]
- Ulusoy I, Yılmaz M, Kıvrak A. How efficient is ChatGPT in accessing accurate and quality health-related information? *Cureus* 2023;15:e46662. [CrossRef]
- Casciato D, Mateen S, Cooperman S, Pesavento D, Brandao RA. Evaluation of online AI-generated foot and ankle surgery information. *J Foot Ankle Surg* 2024;63:680–3. [CrossRef]
- Johnson SB, King AJ, Warner EL, Aneja S, Kann BH, Bylund CL. Using ChatGPT to evaluate cancer myths and misconceptions: Artificial intelligence and cancer information. *JNCI Cancer Spectr* 2023;7:pkad015. [CrossRef]

## ORIJİNAL ÇALIŞMA - ÖZ

### Tetik parmak tedavisinde sık sorulan sorulara ChatGPT'nin yanıtlarının kalitesinin değerlendirilmesi

**AMAÇ:** Bu çalışma, tetik parmak ile ilgili hasta sorularına yanıt vermede Generative Pre-trained Transformer'in (ChatGPT; OpenAI, San Francisco, CA) doğruluk ve güvenilirliğini değerlendirmeyi amaçlamaktadır. Bu değerlendirme, tedavi öncesinde hasta eğitimi geliştirme potansiyeline sahiptir ve yapay zeka tabanlı sistemlerin hasta eğitim sürecindeki rolünü aydınlatmayı hedeflemektedir.

**GEREÇ VE YÖNTEM:** Tetik parmak ile ilgili en sık sorulan on soru, hasta eğitimine yönelik web sitelerinden ve literatür taramasından derlenmiş ve ChatGPT'ye yöneltilmiştir. Yanıtlar, iki ortopedi uzmanı tarafından JAMA Benchmark kriterleri ve DISCERN aracı kullanılarak değerlendirilmiştir. Ek olarak, yanıtların okunabilirliği Flesch-Kincaid sınıf seviyesi ile analiz edilmiştir.

**BULGULAR:** ChatGPT'nin tetik parmak ile ilgili sorulara verdiği yanıtlar için DISCERN puanları 35 ile 47 arasında değişmiş ve ortalama 42 olarak bulunmuştur, bu da "orta" kaliteye işaret etmektedir. Yanıtların %60'ı tatmin edici bulunurken, %40'ında eksiklikler tespit edilmiştir. JAMA Benchmark kriterlerine göre, bilimsel referans eksikliği önemli bir dezavantaj olarak öne çıkmıştır. Ortalama okunabilirlik seviyesi üniversite düzeyindedir, bu da düşük sağlık okuryazarlığına sahip hastalar için bilgiyi anlamayı zorlaştırmaktadır. Yanıtların daha geniş bir hasta kitlesi için erişilebilir ve anlaşılabilir hale getirilmesi gerekmektedir.

**SONUÇ:** Bulgularımız, bildiğimiz kadarıyla, tetik parmak bağlamında ChatGPT kullanımını araştıran ilk çalışma olduğunu göstermektedir. ChatGPT, tetik parmak hakkında genel bilgiler sağlama konusunda makul bir başarı göstermektedir; ancak, hasta eğitimi için birincil kaynak olarak kullanılmadan önce uzman denetimi gereklidir.

**Anahtar sözcükler:** Tetik parmak; ChatGPT; DISCERN; hasta eğitimi; yapay zekâ.

Ulus Travma Acil Cerrahi Derg 2025;31(4):389-393 DOI: 10.14744/tjtes.2025.32735