

Journal of İstanbul Faculty of Medicine İstanbul Tıp Fakültesi Dergisi

Research Article

Open Access

DIGITAL CURIOSITY ABOUT FIBROMYALGIA: AN ANALYSIS OF THE QUALITY AND READABILITY OF ChatGPT'S RESPONSES BASED ON GOOGLE TRENDS

FİBROMİYALJİYE DAİR DİJİTAL MERAK: ChatGPT'NİN GOOGLE TRENDLER TABANLI YANITLARININ KALİTE VE OKUNABİLİRLİK ANALİZİ



Emre Öney¹ , Berk Korkut¹ , Fatma Merih Akpınar¹ , Sina Arman¹ , Demirhan Dıraçoğlu¹ 

¹ İstanbul University, İstanbul Faculty of Medicine, Department of Physical Medicine and Rehabilitation, İstanbul, Türkiye

Abstract

Objective: The aim of this study is to evaluate the quality and readability of patient education materials about fibromyalgia generated by ChatGPT to assess whether this AI application can be reliably used as an online health information source for the general population.

Materials and Methods: In this study, Google Trends was used to identify the most frequently searched keywords related to fibromyalgia. The identified keywords were categorized into three groups: general information, diagnosis, and treatment, and were sequentially entered into ChatGPT. The quality of the generated responses was assessed using the modified DISCERN (mDISCERN) tool. The readability of the texts was evaluated using the Flesch-Kincaid Grade Level (FKGL) and Flesch-Kincaid Reading Ease (FRE) parameters.

Results: The mean mDISCERN score of the texts was 44.31±4.67, the mean FKGL score was 15.29±2.29, and mean FRE score was 24.23±10.27. The average mDISCERN score indicated that the texts were of moderate quality. According to the mean Flesch-Kincaid scores, the texts were found to be quite difficult to read and were mostly appropriate for readers at the undergraduate level or higher. Furthermore, the statistical analysis among the three categories revealed no significant differences in mDISCERN, FKGL, and FKRE scores.

Conclusion: AI language models like ChatGPT hold innovative potential in healthcare communication by providing accessible and unbiased information that can support health literacy and patient education. However, the high level of reading comprehension required currently limits their effectiveness and reliability for direct patient use. Therefore, the content generated by ChatGPT should be considered a complementary resource and should be critically examined by users.

Öz

Amaç: Bu çalışmanın amacı, ChatGPT tarafından fibromiyalji hakkında oluşturulan hasta eğitim materyallerinin kalitesini ve okunabilirliğini değerlendirmek ve bu yapay zeka uygulamasının genel popülasyon için güvenilir bir çevrimiçi sağlık bilgi kaynağı olarak kullanılıp kullanılmayacağına değerlendirmektir.

Gereç ve Yöntemler: Çalışmada, fibromiyalji ile ilgili en sık aranan anahtar kelimeler Google Trendler kullanılarak belirlendi. Belirlenen anahtar kelimeler genel bilgi, tanı ve tedavi olmak üzere üç gruba ayrıldı ve sırayla ChatGPT'ye girildi. Oluşturulan yanıtların kalitesi, modifiye edilmiş DISCERN (mDISCERN) aracı kullanılarak değerlendirildi. Metinlerin okunabilirliği, Flesch-Kincaid Sınıf Düzeyi (FKGL) ve Flesch-Kincaid Okuma Kolaylığı (FRE) parametreleri kullanılarak değerlendirildi.

Bulgular: Metinlerin ortalama mDISCERN puanı 44.31±4.67, ortalama FKGL puanı 15.29±2.29 ve ortalama FRE puanı 24.23±10.27 olarak bulundu. Ortalama mDISCERN puanı, metinlerin orta kalitede olduğunu gösterdi. Ortalama Flesch-Kincaid skorlarına göre, metinlerin oldukça zor okunabilir olduğu ve çoğunlukla lisans düzeyi veya üzeri okuma becerisi gerektirdiği belirlendi. Ayrıca, üç kategori arasındaki istatistiksel analiz, mDISCERN, FKGL ve FKRE puanlarında anlamlı bir fark ortaya koymadı.

Sonuç: Özetle, ChatGPT gibi yapay zekâ dil modelleri, sağlık iletişimde erişilebilir ve tarafsız bilgi sunarak sağlık okuryazarlığını ve hasta eğitimini destekleme açısından yenilikçi bir potansiyele sahiptir. Ancak, yüksek düzeyde okuma-anlama becerisi gerektirmesi, doğrudan hasta kullanımı açısından etkinliğini ve güvenilirliğini sınırlamaktadır. Bu nedenle, ChatGPT tarafından üretilen içerikler tamamlayıcı bir kaynak olarak değerlendirilmelidir ve kullanıcılar tarafından eleştirel biçimde incelenmelidir.



“ Citation: Öney E, Korkut B, Merih Akpınar F, Arman S, Dıraçoğlu D. Digital curiosity about fibromyalgia: an analysis of the quality and readability of chatgpt's responses based on google trends. Journal of İstanbul Faculty of Medicine 2026;89(1):57-62. <https://doi.org/10.26650/IUITFD.1805348>

© This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International License. 

© 2026. Öney E, Korkut B, Merih Akpınar F, Arman S, Dıraçoğlu D.

✉ Corresponding author: Emre Öney emreoney54@hotmail.com



Keywords Fibromyalgia · ChatGPT · Readability · Quality · Artificial intelligence

Anahtar Kelimeler Fibromiyalji, ChatGPT, Okunabilirlik, Kalite, Yapay zeka

INTRODUCTION

Fibromyalgia is a chronic condition characterized by various symptoms, including widespread pain, fatigue, sleep and cognitive disturbances, as well as anxiety, depression, and emotional difficulties. It is reported to affect 2–8% of the global population and is approximately three times more common in women than in men (1).

In treatment approaches, in addition to pharmacotherapy, non-pharmacological methods such as patient education, cognitive behavioral therapy, and increased physical activity are initially recommended (2). In this context, it is critically important for patients to have sufficient knowledge about their condition, as it plays a key role in treatment adherence and quality of life. However, although the internet allows rapid access to health-related information, its reliability as an information source remains limited due to the broad variability in content. In addition, verifying the accuracy of online health information can be challenging for patients (3).

In recent years, AI-powered chatbots have emerged as potential tools for generating health education materials. Through questions directed at these bots, users can access compiled online information. The current literature includes numerous studies examining the quality, readability, and impact of chatbot-generated materials on patient satisfaction (4). For example, ChatGPT's ability to respond in a manner resembling human language patterns, its rapid generation of educational content, and its accessibility are major advantages (5). However, considering that patient education materials are recommended to be written at no higher than a sixth-grade reading level according to standards set by the U.S. Department of Health and Human Services, the American Medical Association, and the National Institutes of Health, AI-generated content still has areas that require improvement in terms of accuracy, consistency, and readability (6). Additionally, issues related to data privacy and ethics must also be addressed prior to implementation.

In the current literature, ChatGPT-4 has been reported to provide “generally good responses” in prostate cancer patient education and may serve as a valuable tool; it was found to prepare kidney transplant brochures rapidly, though with a noted need for greater depth; it approached the target reading level in orthopedic hand surgery materials; and in physical medicine and rehabilitation, it functioned as a virtual educational partner offering support for self-management and motivation (7-10).

Accordingly, the aim of this study is to examine the quality and readability of fibromyalgia-related patient education materials generated by ChatGPT.

MATERIALS AND METHODS

Fibromyalgia-related keywords were identified using Google Trends (<https://trends.google.com/>) on December 15, 2024. The term “fibromyalgia” was analyzed from 2004 to the present to identify the most frequently searched health-related queries directly associated with fibromyalgia. Sixteen keywords that were clearly and specifically related to fibromyalgia were selected from the top 25 most frequently searched terms by prioritizing queries that reflect common patient information needs.

The selected fibromyalgia-related keywords were systematically entered into the free version of ChatGPT (GPT-3.5), which is the most widely accessible and commonly used model by the general population. As the primary aim of this study was to evaluate ChatGPT as a real-world source of patient education and health information, the use of the free version was considered more representative of actual user exposure.

The keywords were classified into three categories: general information, diagnosis, and treatment. Nine keywords (56.3%) were assigned to the general information category, four (25%) to diagnosis, and three (18.7%) to treatment (Table 1).

Table 1. Sixteen fibromyalgia-related keywords based on Google Trends Data

RANK	Keyword	Category
1	Symptoms fibromyalgia	General information
2	Fibromiyalgiya pain	General information
3	Fibromyalgia syndrome	General information
4	Fibromyalgia cause	General information
5	What is fibromyalgia	General information
6	Fibromyalgia treatment	Treatment
7	Fibromyalgia points	Diagnosis
8	Fibromyalgia test	Diagnosis
9	Fibromyalgia fatigue	General information
10	Fibromyalgia diet	Treatment
11	Fibromyalgia back pain	General information
12	Fibromyalgia medication	Treatment
13	Chronic fatigue	General information
14	Fibromyalgia diagnosis	Diagnosis
15	Muscle pain	General information
16	Fibromyalgia autoimmune	Diagnosis



The quality of the generated responses was assessed using the modified DISCERN (mDISCERN) instrument. The evaluation process involved independent review by two authors (EÖ and BK). In cases of discrepancy, a consensus was reached under the guidance of a third reviewer (FMA), who made the final decision.

Similar to the original DISCERN instrument, the modified DISCERN (mDISCERN) scale comprises 16 criteria, each rated on a 5-point Likert scale, resulting in a maximum possible score of 80. The scale evaluates the accuracy and credibility of information, the quality of content related to diagnosis and treatment, and the professional tone of patient education materials (Figure 1) (11). Based on the total mDISCERN score, the material quality was classified into one of the five predefined categories (Table 2) (12).

Questions	
Section 1: Accuracy & Credibility	
1. Are the responses medically accurate?	
2. Does the AI-platform disclose conflicts of interest or declare that it is not software designed to provide accurate medical information?	
3. Does the AI-platform list the sources of information used, including the date?	
4. Can the sources be verified?	
5. Is it balanced or unbiased?	
6. Does it refer to any areas of uncertainty?	
Section 2: Quality of Information	
7. Does it describe the disease, its causes, and symptoms?	
8. Does it describe how the disease can be diagnosed?	
9. Does it describe how the treatment(s) work?	
10. Does it describe the benefits of each treatment?	
11. Does it describe the risks of each treatment?	
12. Does it describe what would happen if no treatment is used?	
13. Is it clear that there may be more than one possible treatment choice?	
14. Does it recommend seeing a healthcare provider if experiencing this disease?	
15. Is it free from major errors that may cause harm?	
Section 3: Professionalism and Tone	
16. Does the AI use a professional, neutral tone?	

Figure 1. mDISCERN Scale

Table 2. mDISCERN scale score

Total mDISCERN score	Quality
16-26	Very poor
27-38	Poor
39-50	Moderate
51-62	Good
63-80	Excellent

The readability of the ChatGPT-generated responses was analyzed using the Flesch-Kincaid Grade Level (FKGL) and Flesch-Kincaid Reading Ease (FRE) metrics (Table 3) (13, 14). The FRE score estimates the difficulty of reading a text on a scale from 0 to 100, with lower scores indicating more difficult texts. FKGL is a widely accepted readability formula that estimates the educational grade level required to comprehend a given text, based on average sentence

length and word complexity. FKGL reflects the level of education needed to understand the material (15).

Table 3. Calculation formulas for flesch-kincaid reading ease and grade level

Readability score	Formula
Flesch reading ease (FRE)	$206.835 - 1.015 \times (\text{Total words}/\text{Total sentences}) - 84.6 \times (\text{Total syllables}/\text{Total words})$
Flesch kincaid grade level (FKGL)	$0.39 \times (\text{Total words}/\text{Total sentences}) + 11.8 \times (\text{Total syllables}/\text{Total words}) - 15.59$

The mDISCERN, FKGL, and FRE scores of ChatGPT-generated texts for each category (general information, diagnosis, and treatment) were presented using descriptive statistics (mean, standard deviation, minimum and maximum). The Kruskal-Wallis test was performed to assess whether there were statistically significant differences among the three categories in terms of these scale scores. The significance level was set at $p < 0.05$. Statistical analyses were conducted using IBM SPSS Statistics, Version 21.0 (IBM Corp., Armonk, NY, USA).

RESULTS

The most frequently searched keyword on Google Trends was symptoms fibromyalgia.

The mean mDISCERN score of the texts was 44.31 ± 4.67 , the mean FKGL score was 15.29 ± 2.29 , and the mean FRE score was 24.23 ± 10.27 (Table 4). The average mDISCERN score indicated that the texts were of moderate quality. According to the mean Flesch-Kincaid scores, the texts were found to be quite difficult to read and were mostly appropriate for readers at the undergraduate level or higher.

Table 4. Statistical summary of the scales

Scales	n	Minimum	Maximum	Mean	Standard deviation
mDISCERN	16	36.00	51.00	44.31	4.67
Flesch kincaid grade level	16	12.08	20.95	15.29	2.28
Flesch kincaid reading ease	16	1.31	42.61	24.23	10.27

The mean total mDISCERN scores were 44.67 ± 5.36 for the general information category, 41.50 ± 3.11 for the diagnosis category, and 47.00 ± 2.65 for the treatment category. No statistically significant difference was found among the three categories in terms of mDISCERN scores ($p = 0.365$). The FKGL scores were 14.33 ± 1.54 for general information, 17.50 ± 3.04 for diagnosis, and 15.24 ± 1.47 for treatment. No statistically significant difference was observed between groups regarding FKGL values ($p = 0.146$). The FRE scores were 27.65 ± 8.95 for the general information category, 18.44 ± 13.14 for diagnosis, and 21.65 ± 9.04 for treatment. No significant difference was detected between categories regarding FRE scores ($p = 0.399$) (Table 5).

Table 5. Comparison of mean scores for mDISCERN, Flesch-Kincaid Reading Ease, and Flesch-Kincaid Grade Level

Scales	n	Minimum	Maximum	Mean	Standard deviation	p*
mDISCERN						
General information	9	36.00	51.00	44.67	5.36	0.365
Diagnosis	4	37.00	44.00	41.50	3.11	
Treatment	3	44.00	49.00	47.00	2.65	
Flesch kincaid grade level						
General information	9	12.08	17.36	14.33	1.54	0.146
Diagnosis	4	14.23	20.95	17.50	3.04	
Treatment	3	13.57	16.31	15.24	1.47	
Flesch kincaid reading ease						
General information	9	13.54	42.61	27.65	8.95	0.399
Diagnosis	4	1.31	31.53	18.44	13.14	
Treatment	3	16.31	32.09	21.65	9.04	

*Kruskal Wallis Test

DISCUSSION

This study examined the quality and readability of texts about fibromyalgia generated by ChatGPT, highlighting the potential and limitations of this technology as a health information resource. While ChatGPT provides easily accessible information, the quality of these materials was found to be moderate. Moreover, the fibromyalgia-related texts produced by ChatGPT were evaluated as lacking readability even for individuals at the high school education level.

The moderate quality level of ChatGPT's responses concerning fibromyalgia indicates that the answers are informative yet only partially reliable. The responses generally provide a broad overview of the disease's symptoms and treatment options. However, deficiencies such as the absence of source citations, lack of references to current evidence-based guidelines, and frequent omission of personalized medical advice detract from the overall quality of the responses.

In our study, the 16 keywords selected as directly related to fibromyalgia were intended to reflect patients' common information needs; however, the relatively limited number of keywords may influence the robustness and generalizability of the findings. The literature indicates that in AI-based health information studies, analyses are often conducted using a limited set of high-search volume keywords. While this approach offers advantages in representing typical user queries, the use of broader keyword sets may lead to variability in the results. Nevertheless, the selection of keywords based on Google Trends data and their explicit relevance to fibromyalgia indicates that the findings capture the main underlying

trends. Future studies expanding the number of analyzed keywords may help test the findings across different contexts and strengthen their generalizability.

Bias and ethical considerations represent important aspects of AI-generated health information. Although bias was not directly assessed in the present study, previous literature has highlighted that AI models trained on large-scale datasets may reflect and reproduce existing cultural, political, and ideological biases embedded within those data. As a result, AI-generated content may vary across different religious beliefs, ethnic identities, or languages. For instance, Huang et al. reported that demographic biases are prevalent in many clinical machine learning models and may have implications for model performance and patient safety (16). Similarly, Yang et al. demonstrated that large language models may generate more detailed and positive narratives for white patients while producing shorter and less specific content for Black and Latino patients in medical reporting contexts (17). These findings underscore broader ethical considerations relevant to AI-based health information systems and highlight the need for a cautious interpretation of AI-generated content.

The indicators in our study demonstrate that the health information texts generated by ChatGPT require an advanced level of reading ability, which makes them difficult for the general population to read. In contrast, many health organizations recommend that patient education materials be written at a 6th to 8th grade reading level (18, 19). Previous studies have reported that decreasing readability levels may hinder the comprehension of health information, and these findings support the results of our study (20, 21). Nevertheless, clinical outcomes such as patient comprehension, behavioral responses, or treatment adherence were not directly assessed in the present study. Additionally, the extensive use of medical terminology and the preference for an academic writing style in some generated texts may further compromise their comprehensibility. Although Ayre et al. demonstrated that ChatGPT has the potential to simplify health-related texts, they reported that even simplified content does not consistently fall below the recommended readability levels; this observation is consistent with the findings of our study (22). In this context, the potential clinical implications of reduced readability should be interpreted cautiously considering in light of the existing literature, rather than as direct outcomes of the present study.

These findings highlight the necessity for improvements in both content quality and language simplification in AI-based text generation. Firstly, the accuracy and currency of information provided by ChatGPT should be validated through expert review; to mitigate the risk of misleading or incomplete information, all texts must undergo supervision by healthcare professionals. Secondly, to enhance health literacy and broaden access among diverse patient populations, readability levels should be lowered and adapted



to more comprehensible standards by applying natural language processing techniques and text revision strategies.

The limitations of this study include the evaluation being conducted using a single artificial intelligence model and the analyses being limited to 16 English language keywords. In addition, the exclusive use of the free version of ChatGPT (GPT-3.5) represents an important limitation when considering the rapid evolution of large language models. Indeed, more advanced versions, such as GPT-4, have been reported to demonstrate superior performance in terms of accuracy, contextual understanding, and citation capabilities. Therefore, the findings of the present study may not be directly generalizable to newer or paid versions of ChatGPT, and potential differences in their performance within patient education contexts should be considered. Future studies comparing different large language models, expanding the number of analyzed keywords, conducting multilingual analyses, and performing subgroup evaluations across diverse patient populations (e.g., age, education level, and cultural background) may help improve the comprehensiveness and generalizability of the findings.



Peer Review Externally peer-reviewed.

Author Contributions Conception/Design of Study- E.Ö., B.K., M.A. Data Acquisition- E.Ö., B.K., M.A. Data Analysis/Interpretation- E.Ö., B.K., M.A., S.A., D.D.; Drafting Manuscript- E.Ö., B.K., M.A., S.A., D.D.; Critical Revision of Manuscript- E.Ö., B.K., M.A., S.A., D.D.; Final Approval and Accountability- E.Ö., B.K., M.A., S.A., D.D..

Conflict of Interest The authors declared no conflict of interest.

Financial Disclosure The authors declared that they received no financial support for this study.

Author Details

Emre Öney

¹ İstanbul University, İstanbul Faculty of Medicine, Department of Physical Medicine and Rehabilitation, İstanbul, Türkiye

0009-0001-8611-8282 emreoney54@hotmail.com

Berk Korkut

¹ İstanbul University, İstanbul Faculty of Medicine, Department of Physical Medicine and Rehabilitation, İstanbul, Türkiye

0009-0004-6621-8237

Fatma Merih Akpınar

¹ İstanbul University, İstanbul Faculty of Medicine, Department of Physical Medicine and Rehabilitation, İstanbul, Türkiye

0000-0001-8072-5201

Sina Arman

¹ İstanbul University, İstanbul Faculty of Medicine, Department of Physical Medicine and Rehabilitation, İstanbul, Türkiye

0000-0003-4885-389X

CONCLUSION

In conclusion, AI language models, such as ChatGPT, hold potential as innovative tools in health communication. Based on the obtained data and existing literature, ChatGPT offers several advantages, including accessibility and neutrality, which may theoretically support health literacy and patient education. However, its high reading level requirements and shortcomings in source citation necessitate cautious consideration regarding patient education and safety in its current form. Therefore, ChatGPT outputs should be regarded solely as supplementary information sources and critically appraised by patients. Future advancements with more sophisticated models and specialized training strategies may help overcome these limitations, enabling a more effective use of AI in health literacy and patient education.

Demirhan Dıraçoğlu

¹ İstanbul University, İstanbul Faculty of Medicine, Department of Physical Medicine and Rehabilitation, İstanbul, Türkiye

0000-0001-5682-5050

REFERENCES

- Filipovic T, Filipović A, Nikolic, Gimigliano F, Stevanov J, Bosanac I. Fibromyalgia: Understanding, diagnosis and modern approaches to treatment. *J Clin Med* 2025;14(3):955.
- Goldenberg DL, Kaplan M. Fibromyalgia: Treatment in adults. In: Connor RF, editor. *UpToDate*. Wolters Kluwer; 2025.
- Golan R, Reddy R, Ramasamy R. The rise of artificial intelligence-driven health communication. *Transl Androl Urol* 2024;13(2):356-8.
- Behers BJ, Stephenson-Moe CA, Gibbons RM, Vargas IA, Wojtas CN, Rosario MA, et al. Assessing the quality of patient education materials on cardiac catheterization from artificial intelligence chatbots: An observational cross-sectional study. *Cureus* 2024;16(9):e69996.
- Günay AE, Özer A, Yazıcı A, Sayer G. Comparison of ChatGPT versions in informing patients with rotator cuff injuries. *JSES Int* 2024;8(5):1016-8.
- Ozduran E, Hancı V, Erkin Y, Özbek İC, Abdulkirimov V. Assessing the readability, quality and reliability of responses produced by ChatGPT, Gemini, and Perplexity regarding most frequently asked keywords about low back pain. *PeerJ* 2025;13:e18847.
- Gibson D, Jackson S, Shanmugasundaram R, Seth I, Siu A, Ahmadi N, et al. Evaluating the efficacy of ChatGPT as a patient education tool in prostate cancer: Multimetric assessment. *J Med Internet Res* 2024;26:e55939.
- Tian Tran J, Burghall A, Blydt-Hansen T, Cammer A, Goldberg A, Hamiwa L, et al. Exploring the ability of ChatGPT to create quality patient education resources about kidney transplant. *Patient Educ Couns* 2024;129:108400.
- Vallurupalli M, Shah ND, Vyas RM. Optimizing readability of patient-facing hand surgery education materials using Chat Generative Pretrained Transformer (ChatGPT) 3.5. *J Hand Surg Am* 2024;49(10):986-91.
- Peng S, Wang D, Liang Y, Xiao W, Zhang Y, Liu L. AI-ChatGPT/GPT-4: An booster for the development of physical medicine and rehabilitation in the new era. *Ann Biomed Eng* 2024;52(3):462-6.



- 11 Ghanem YK, Rouhi AD, Al-Houssan A, Saleh Z, Moccia MC, Joshi H, et al. Dr. Google to Dr. ChatGPT: Assessing the content and quality of artificial intelligence-generated medical information on appendicitis. *Surg Endosc* 2024;38(5):2887-93.
- 12 D'Ambrosi R, Annibaldi A, Carrozzo A, Vieira TD, An JS, Freychet B, et al. Evaluating the reliability of YouTube as a source of information for meniscal ramp lesions. *Orthop J Sports Med* 2024;12(1):23259671231219815.
- 13 Flesch R. A new readability yardstick. *J Appl Psychol* 1948;32(3):221-33.
- 14 Kincaid JP, Fishburne RP Jr, Rogers RL, Chissom BS. Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel. *Res Branch Rep* 1975;8-75.
- 15 Guven Y, Ozdemir OT, Kavan MY. Performance of artificial intelligence chatbots in responding to patient queries related to traumatic dental injuries: A comparative study. *Dent Traumatol* 2025;41(3):338-47.
- 16 Huang J, Galal G, Etemadi M, Vaidyanathan M. Evaluation and mitigation of racial bias in clinical machine learning models: Scoping review. *JMIR Med Inform* 2022;10(5):e36388.
- 17 Yang Y, Liu X, Jin Q, Huang F, Lu Z. Unmasking and quantifying racial bias of large language models in medical report generation. *Commun Med* 2024;4(1):176.
- 18 Rooney MK, Santiago G, Perni S, Horowitz DP, McCall AR, Einstein AJ, et al. Readability of patient education materials from high-impact medical journals: A 20-year analysis. *J Patient Exp* 2021;8:2374373521998847.
- 19 Badarudeen S, Sabharwal S. Assessing readability of patient education materials: Current role in orthopaedics. *Clin Orthop Relat Res* 2010;468(10):2572-80.
- 20 Özbek İC, Hancı V, Özduran E. Digital Guidance: Quality and readability analysis of artificial intelligence-generated spondyloarthropathy texts. *Turk J Osteoporos* 2025;31(1):12-18.
- 21 Rouhi AD, Ghanem YK, Yolchieva L, et al. Can Artificial intelligence improve the readability of patient education materials on aortic stenosis? A Pilot Study. *Cardiol Ther* 2024;13(1):137-147.
- 22 Ayre J, Mac O, McCaffery K, McKay BR, Liu M, Shi Y, et al. New frontiers in health literacy: Using ChatGPT to simplify health information for people in the community. *J Gen Intern Med* 2024;39(4):573-7.

