

Can Generative Artificial Intelligence Enhance Health Literacy About Lateral Epicondylitis?

Review began 05/21/2024
Review ended 05/27/2024
Published 05/30/2024

© Copyright 2024
Miskiewicz et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Michael J. Miskiewicz¹, Christian Leonardo¹, Salvatore Capotosto¹, Kenny Ling², Dorian Cohen¹, David Komatsu¹, Edward D. Wang¹

1. Department of Orthopaedic Surgery, Stony Brook University, Stony Brook, USA 2. Department of Orthopaedics, Stony Brook University Hospital, Stony Brook, USA

Corresponding author: Michael J. Miskiewicz, michael.miskiewicz@stonybrookmedicine.edu

Abstract

Introduction: Health literacy is a critical determinant of a patient's overall health status, and studies have demonstrated a consistent link between poor health literacy and negative health outcomes. The Centers for Disease Control and Prevention (CDC) and the National Institutes of Health (NIH) advise that patient educational materials (PEMs) should be written at an eighth-grade reading level or lower, matching the average reading level of adult Americans. The purpose of this study was to investigate the ability of generative artificial intelligence (AI) to edit PEMs from orthopaedic institutions to meet the CDC and NIH guidelines.

Methods: PEMs about lateral epicondylitis (LE) from the top 25 ranked orthopaedic institutions from the 2022 U.S. News & World Report Best Hospitals Specialty Ranking were gathered. ChatGPT Plus (version 4.0) was then instructed to rewrite PEMs on LE from these institutions to comply with CDC and NIH-recommended guidelines. Readability scores were calculated for the original and rewritten PEMs, and paired t-tests were used to determine statistical significance.

Results: Analysis of the original and edited PEMs about LE revealed significant reductions in reading grade level and word count of 3.70 ± 1.84 ($p < 0.001$) and 346.72 ± 364.63 ($p < 0.001$), respectively.

Conclusion: Our study demonstrated generative AI's ability to rewrite PEM about LE at a reading comprehension level that conforms to the CDC and NIH guidelines. Hospital administrators and orthopaedic surgeons should consider the findings of this study and the potential utility of artificial intelligence when crafting PEMs of their own.

Categories: Medical Education, Orthopedics, Healthcare Technology

Keywords: health literacy, patient education materials, chatgpt, lateral epicondylitis, orthopedic surgery

Introduction

Health literacy, or the ability of patients to understand health-related information, stands as one of the most significant indicators of an individual's overall health [1]. Past research has consistently linked low health literacy to negative health outcomes [2-3]. The surge in global internet usage by over 11-fold in recent years [3] presents healthcare providers with a new opportunity to disseminate information to their patients. This shift towards online patient education necessitates an awareness of the average patient's reading comprehension abilities by physicians to ensure the highest quality of patient education material (PEM).

Data from the National Center for Education Statistics indicate that in the United States, the average person reads at about the level of an eighth grader [4]. Considering this, both the Centers for Disease Control and Prevention (CDC) and the National Institute of Health (NIH) advise that PEMs should be written at no higher than an eighth-grade reading level [5-6]. However, research indicates that many hospital and physician websites offer PEMs at reading levels that far exceed these recommendations [7-10]. This situation underscores the pressing need for PEMs to be revised to a level accessible and understandable to the general population.

Lateral epicondylitis (LE), commonly known as "tennis elbow," represents a form of tendinosis of the extensor muscles of the forearm and is one of the most common sources of lateral elbow pain [11]. This condition has traditionally been linked to physical activities that require repetitive motion of the arm and wrist, such as playing tennis or using a screwdriver at work [12]. Consequently, blue-collar workers whose jobs predominantly involve manual labour may be at particularly high risk for developing LE [11-13]. This raises concerns about the accessibility and effectiveness of PEMs targeted to individuals suffering from LE.

One potential solution to this discrepancy between the CDC's recommendations and the actual readability of available online PEMs involves utilising artificial intelligence. In response to the skyrocketing global interest in artificial intelligence (AI), many companies are exploring innovative ways to introduce AI-based

How to cite this article

Miskiewicz M J, Leonardo C, Capotosto S, et al. (May 30, 2024) Can Generative Artificial Intelligence Enhance Health Literacy About Lateral Epicondylitis?. Cureus 16(5): e61384. DOI 10.7759/cureus.61384

technology into everyday life. One notable example of such companies is OpenAI with its widely used online chatbot, Chat Generative Pretrained Transformer (ChatGPT). ChatGPT is a natural language processing tool that engages users in a conversational dialogue [14]. While the full versatility of ChatGPT is still being explored, it attracts over one hundred million weekly users who utilise it for professional, educational, and recreational purposes [14-16].

The goal of this study is to explore the ability of ChatGPT to generate PEMs on LE that are comprehensible to most Americans, while considering the average national reading grade level. We hypothesise that by incorporating recommendations proposed by the CDC and NIH, ChatGPT will be able to improve the readability of these PEMs in efforts to bridge the gap in health literacy between patient and provider.

Materials And Methods

The top 25 orthopaedic institutions, as ranked by the 2022 U.S. News & World Report Best Hospitals Specialty Ranking, were selected for this study. The websites of these institutions were searched for PEMs pertaining to LE. The content from each website was copied and preserved for analysis, excluding any audiovisual multimedia such as pictures, diagrams, and videos. Institutions lacking relevant PEMs on LE were excluded from the study.

Each institution’s PEMs were individually uploaded to ChatGPT Plus (version 4.0). ChatGPT was prompted with a message to rewrite the PEM with the following parameters: (1) limit the total number of polysyllabic words to less than 30, (2) limit sentences to less than 10 words, (3) limit paragraphs to less than five sentences, (4) eliminate as much medical jargon without compromising accuracy, (5) when eliminating medical jargon is not possible, provide a brief explanation of the relevant concept, and (6) overall, rewrite this as if you were speaking to an eighth grader. These parameters were based on pertinent recommendations set forth by both the CDC’s Simply Put and NIH Clear & Simple documents [5-6]. A senior orthopaedic surgery resident reviewed all rewritten PEMs to verify the accuracy of the information presented.

Readability scores for both the original PEMs and those rewritten by ChatGPT were calculated using the Readability Formulas website [17]. For each set of PEMs, readability was assessed using seven distinct formulas: Gunning Fog, Flesch-Kincaid Grade Level, Coleman-Liau Index, Simple Measure of Gobbledygook (SMOG) Index, Automated Readability Index, Linsear Write Formula, and FORECAST Readability Formula. These formulas evaluate different elements of text composition, including total word count, the presence of polysyllabic words, and overall language complexity, to derive a readability score, which is indicative of the text’s grade-level comprehension requirement (Table 1). The study calculated the average readability score and word count for the PEMs of each institution. To assess the significance of the differences between the original and ChatGPT-revised PEMs, paired t-tests were conducted. Statistical analyses were performed using SPSS software (Version 29.0.0(241); Armonk, NY: IBM Corp), with a p-value of 0.05 or lower indicating statistical significance.

Readability Tool	Formula
Gunning Fog	Grade level = 0.4 (average sentence length/percentage of hard words)
Flesch-Kincaid Grade Level	Grade level = (0.39 x average sentence length) + (11.8 x average # syllable per word) - 15.59
The Coleman-Liau Index	Grade level = 0.0855 (average # of letters per 100 words) - 0.296 (average # of sentences per 100 words) - 15.8
SMOG Index	Grade level = 3 x square root of polysyllable count
Automated Readability Index	Grade level = 4.71 (characters/words) + 0.5 (words/sentences) - 21.43
Linsear Write Formula	n = [(2 syllable words x 1) + (3 or more syllable words x 3)]/number of sentences If n <20, Grade level = n/20 If n >20, Grade level = n-2/20
FORCAST Readability Formula	Grade level = 20 - (# of single syllable words x 150 / # of words x 10)

TABLE 1: Description of readability tools and their corresponding formulas.

Results

Twenty-two of the initial 25 orthopaedic institutions contained educational material related to LE on their

websites. In the original, unedited PEM cohort, only six institutions obtained average readability scores below the eighth-grade reading level. The average reading grade level of all institutions' original PEMs was 9.81 ± 1.76 , and the average word count was 600.68 ± 409.44 words. Following ChatGPT's edits to the original PEMs, all 25 rewritten PEMs obtained average readability scores below the eighth-grade reading level. The total average readability score for all rewritten PEMs was 6.12 ± 0.97 (Figure 1), with an average word count of 253.96 ± 100.76 (Figure 2). By utilising ChatGPT to rewrite the original PEMs, a reduction of 3.70 ± 1.84 ($p < 0.001$) reading grade levels and 346.72 ± 364.63 ($p < 0.001$) words was achieved (Table 2). The senior orthopaedic resident validated the accuracy of the information in the PEMs rewritten by ChatGPT.

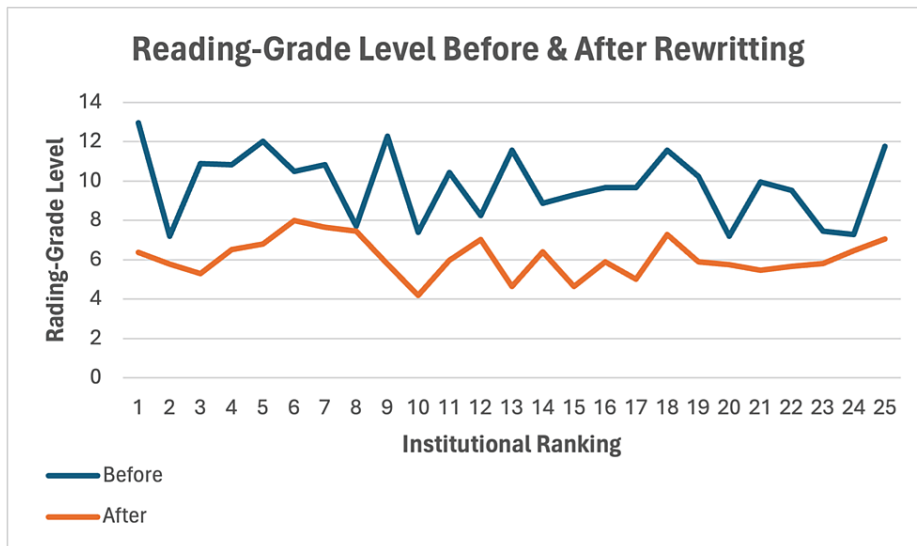


FIGURE 1: The average readability scores of patient education materials related to lateral epicondylitis from 25 of the top nationally ranked orthopaedic institutions, before and after generative AI-assisted editing.

AI: artificial intelligence.

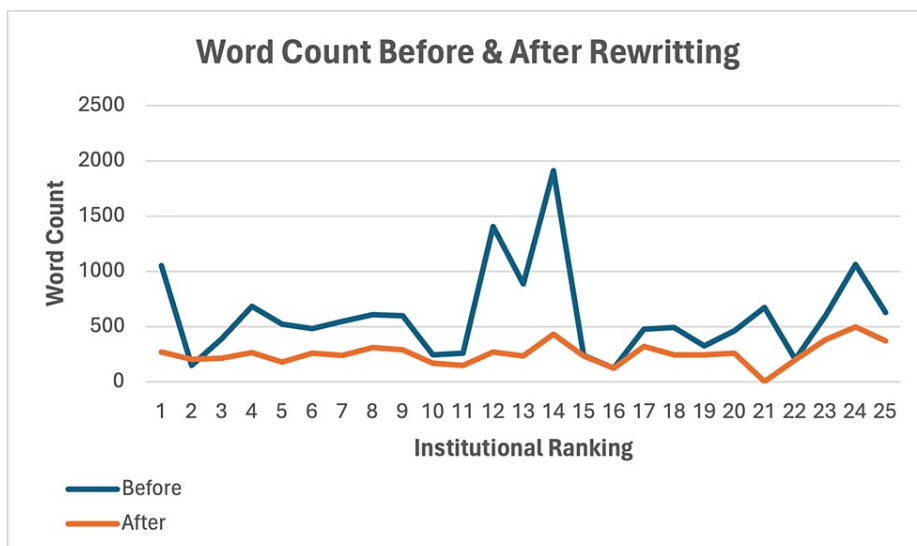


FIGURE 2: The average word count of patient education materials related to lateral epicondylitis from 25 of the top nationally ranked orthopaedic institutions, before and after generative AI-assisted editing.

AI: artificial intelligence.

	Mean Reduction	Standard Deviation	95% Confidence Interval		p-value
			Lower	Upper	
Readability score	3.70	1.84	2.93	4.46	<0.001
Word count	346.72	364.63	196.21	497.24	<0.001

TABLE 2: Results for paired t-test.

Results for paired t-test to determine significance between (a) average readability scores and (b) total word count of the original and rewritten PEMs. PEM: patient educational material.

Discussion

Our study examined patient educational material provided by 25 distinct orthopaedic institutions. Among them, 22 included PEMs pertaining to lateral epicondylitis. Only six of these institutions initially had PEM written at an eighth-grade reading level or lower. With minimal guidance, ChatGPT successfully revised the original PEMs, resulting in a decrease in average reading grade level and word count to meet the recommended guidelines by the CDC and NIH. These findings suggest that ChatGPT shows promise in enhancing the comprehensibility of PEMs for their target audience.

With the widespread availability of the internet at home, work, and while on the move, accessing PEMs have become more convenient than ever. Despite this improved access, the understanding of these materials has not seen a corresponding increase. Previous studies have shown a significant discrepancy between the CDC and NIH-recommended guidelines for writing PEMs and the true readability of PEMs published by healthcare institutions [18-20]. The implications of these studies are profound, as there is a well-documented correlation between low health literacy and worse clinical outcomes [1-3,21]. Moreover, a study investigating the impact of health literacy on patient satisfaction in surgical settings revealed that patients with higher health literacy tended to report greater satisfaction with their surgeries compared to those with lower health literacy level [22]. Health literacy has also been shown to be positively correlated with patient compliance [23]. Overall, PEMs that adhere to CDC and NIH guidelines not only improve patient outcomes, satisfaction, and compliance but also empower patients to make informed decisions about their healthcare. Artificial intelligence may be the solution to writing simpler and more effective PEMs.

In recent decades, there has been significant progress in artificial intelligence, with ChatGPT leading the way. These advancements have led to ChatGPT's widespread integration across various sectors, including the medical field. A recent study on ChatGPT's ability to develop customized obesity treatment plans revealed its proficiency in crafting individualized strategies tailored to the specific requirements of each person [24]. Another study that looked at ChatGPT's ability to manage obstructive sleep apnoea (OSA) determined that ChatGPT was able to demonstrate potential as a valuable resource for OSA diagnoses [25]. These studies showcase the multifaceted utility of ChatGPT within the healthcare sector, displaying its potential to revolutionise personalized treatment strategies for conditions like obesity and OSA. Building on this foundation, our study indicates that ChatGPT's capabilities extend further, demonstrating its capacity to craft PEMs that abide by the standards given by the CDC and NIH. This not a standalone finding. Previous studies have found that ChatGPT is not only capable of generating text-based information for specific audiences but is also preferred over content generated by humans [26-27]. This not only highlights ChatGPT's versatility but also underscores its pivotal role in enhancing patient comprehension and empowerment in medical decision-making processes [28-29].

There are a few noteworthy limitations to the present study. First, the CDC guidelines suggest considering demographic factors such as race, gender, and ethnicity when writing PEMs, which can be challenging when requesting assistance from a language processing model such as ChatGPT. Additionally, our evaluation of reading comprehension relied solely on reading grade-level comprehension scores and excluded videos, pictures, or other visual media that were included on institution's websites. To our knowledge, there is no reliable metric that would have allowed us to objectively compare audiovisual media between hospital websites. Furthermore, ChatGPT is limited in its ability to incorporate, analyse, and regenerate pictures or diagrams, which was a critical aspect of our methodology. Ultimately, these visual tools may alter and potentially improve the viewer's understanding of the information provided. Potential avenues for future research may include using AI to generate improved audiovisual multimedia in the healthcare space. Despite these limitations, our study is the first to successfully explore ChatGPT's ability to generate PEMs that adhere to the CDC and NIH guidelines.

Conclusions

Our study determined that ChatGPT was capable of rewriting PEM on LE at a reading comprehension level

that abides by the CDC and NIH guidelines. Hospital administrators and orthopaedic surgeons should consider the findings of this study and the potential utility of artificial intelligence when crafting PEMs of their own.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Michael J. Miskiewicz

Acquisition, analysis, or interpretation of data: Michael J. Miskiewicz, Salvatore Capotosto, Christian Leonardo, Kenny Ling, Dorian Cohen, David Komatsu, Edward D. Wang

Drafting of the manuscript: Michael J. Miskiewicz, Salvatore Capotosto, Christian Leonardo

Critical review of the manuscript for important intellectual content: Michael J. Miskiewicz, Salvatore Capotosto, Kenny Ling, Dorian Cohen, David Komatsu, Edward D. Wang

Supervision: Michael J. Miskiewicz, Christian Leonardo, David Komatsu, Edward D. Wang

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue.

Animal subjects: All authors have confirmed that this study did not involve animal subjects or tissue.

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Arslan S: Exploring the potential of Chat GPT in personalized obesity treatment . *Ann Biomed Eng.* 2023, 51:1887-8. [10.1007/s10439-023-05227-9](https://doi.org/10.1007/s10439-023-05227-9)
2. Baskaradoss JK: Relationship between oral health literacy and oral health status . *BMC Oral Health.* 2018, 18:172. [10.1186/s12903-018-0640-1](https://doi.org/10.1186/s12903-018-0640-1)
3. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K: Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med.* 2011, 155:97-107. [10.7326/0003-4819-155-2-201107190-00005](https://doi.org/10.7326/0003-4819-155-2-201107190-00005)
4. Broderick JM, McCarthy A, Hogan N: Osteotomy around the knee: assessment of quality, content and readability of online information. *Knee.* 2021, 28:139-50. [10.1016/j.knee.2020.11.010](https://doi.org/10.1016/j.knee.2020.11.010)
5. Buchanan BK, Varacallo M: Lateral epicondylitis (tennis elbow). *StatPearls [Internet]. StatPearls Publishing, Treasure Island (FL); 2023.*
6. Centers for Disease Control and Prevention, U.S. Department of Health and Human Services: Simply Put: a guide for creating easy-to-understand materials. (2010). Accessed: January 10, 2023: https://www.cdc.gov/healthliteracy/pdf/Simply_Put.pdf.
7. Choudhery S, Xi Y, Chen H, et al.: Readability and quality of online patient education material on websites of Breast Imaging Centers. *J Am Coll Radiol.* 2020, 17:1245-51. [10.1016/j.jacr.2020.04.016](https://doi.org/10.1016/j.jacr.2020.04.016)
8. Ekkel E, Seeras K: Readability of online patient material provided by Reflux Centers in the United States . *Am Surg.* 2023, 89:2782-4. [10.1177/00031348211050828](https://doi.org/10.1177/00031348211050828)
9. OpenAI: GPT-4 technical report. (2024). Accessed: February 8, 2024: <https://cdn.openai.com/papers/gpt-4.pdf>.
10. Hartnett DA, Philips AP, Daniels AH, Blankenhorn BD: Readability of online foot and ankle surgery patient education materials. *Foot Ankle Spec.* 2022, 19586400221116463. [10.1177/19586400221116463](https://doi.org/10.1177/19586400221116463)
11. Herquelot E, Bodin J, Roquelaure Y, et al.: Work-related risk factors for lateral epicondylitis and other cause of elbow pain in the working population. *Am J Ind Med.* 2013, 56:400-9. [10.1002/ajim.22140](https://doi.org/10.1002/ajim.22140)
12. Lenoir H, Mares O, Carlier Y: Management of lateral epicondylitis . *Orthop Traumatol Surg Res.* 2019, 105:S241-6. [10.1016/j.otsr.2019.09.004](https://doi.org/10.1016/j.otsr.2019.09.004)
13. Lockie E, Choi J: Evaluation of a Chat GPT generated patient information leaflet about laparoscopic cholecystectomy. *ANZ J Surg.* 2024, 94:353-5. [10.1111/ans.18834](https://doi.org/10.1111/ans.18834)
14. Martin CA, Khan S, Lee R, Do AT, Sridhar J, Crowell EL, Bowden EC: Readability and suitability of online patient education materials for glaucoma. *Ophthalmol Glaucoma.* 2022, 5:525-30. [10.1016/j.ogla.2022.03.004](https://doi.org/10.1016/j.ogla.2022.03.004)
15. Miller TA: Health literacy and adherence to medical treatment in chronic and acute illness: a meta-analysis . *Patient Educ Couns.* 2016, 99:1079-86. [10.1016/j.pec.2016.01.020](https://doi.org/10.1016/j.pec.2016.01.020)
16. Mira FA, Favier V, Dos Santos Sobreira Nunes H, et al.: Chat GPT for the management of obstructive sleep apnea: do we have a polar star?. *Eur Arch Otorhinolaryngol.* 2024, 281:2087-93. [10.1007/s00405-023-08270-9](https://doi.org/10.1007/s00405-023-08270-9)

17. Miskiewicz M, Capotosto S, Wang ED: Evaluation of readability of patient education materials on lateral epicondylitis (tennis elbow) from the top 25 orthopedic institutions. *JSES Int.* 2023, 7:877-80. [10.1016/j.jseint.2023.05.006](https://doi.org/10.1016/j.jseint.2023.05.006)
18. Mondal H, Mondal S, Podder I: Using ChatGPT for writing articles for patients' education for dermatological diseases: a pilot study. *Indian Dermatol Online J.* 2023, 14:482-6. [10.4103/idoj.idoj_72_23](https://doi.org/10.4103/idoj.idoj_72_23)
19. Mondal H, Panigrahi M, Mishra B, Behera JK, Mondal S: A pilot study on the capability of artificial intelligence in preparation of patients' educational materials for Indian public health issues. *J Family Med Prim Care.* 2023, 12:1659-62. [10.4103/jfmprc.jfmprc_262_23](https://doi.org/10.4103/jfmprc.jfmprc_262_23)
20. National Center for Education Statistics: Digest of education statistics: 2021. (2021). Accessed: February 8, 2024: <https://nces.ed.gov/programs/digest/d21/>.
21. National Institutes of Health: Clear & Simple: developing effective print materials for low-literate readers . (1994). Accessed: November 7, 2022: <https://www.nih.gov/institutes-nih/nih-office-director/office-communications-public-liaison/clear-communication/clear....>
22. Ó Doinn T, Broderick JM, Clarke R, Hogan N: Readability of patient educational materials in sports medicine. *Orthop J Sports Med.* 2022, 10:23259671221092356. [10.1177/23259671221092356](https://doi.org/10.1177/23259671221092356)
23. OpenAI: Introducing ChatGPT. (2024). Accessed: February 8, 2024: <https://openai.com/index/chatgpt/>.
24. Paasche-Orlow MK, Wolf MS: The causal pathways linking health literacy to health outcomes. *Am J Health Behav.* 2007, 31:S19-26.
25. Readability Formulas: Readability Scoring System. (2024). Accessed: February 8, 2024: <https://readabilityformulas.com/readability-scoring-system.php#formulaResults>.
26. White S, Chen J, Atchison R: Relationship of preventive health practices and health literacy: a national study. *Am J Health Behav.* 2008, 32:227-42.
27. Yim CK, Shumate L, Barnett SH, Leitman IM: Health literacy assessment and patient satisfaction in surgical practice. *Ann Med Surg (Lond).* 2018, 35:25-8. [10.1016/j.amsu.2018.08.022](https://doi.org/10.1016/j.amsu.2018.08.022)
28. Reddy S: Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implement Sci.* 2024, 19:27. [10.1186/s13012-024-01357-9](https://doi.org/10.1186/s13012-024-01357-9)
29. Alowais SA, Alghamdi SS, Alsuebany N, et al.: Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ.* 2023, 23:689. [10.1186/s12909-023-04698-z](https://doi.org/10.1186/s12909-023-04698-z)