AKDENİZ SAĞLIK BİLİMLERİ
GÜNCEL ARAŞTIRMALAR DERGİSİ
MEDITERRANEAN JOURNAL OF
HEALTH SCIENCES AND CURRENT RESEARCH
OPEN ACCESS

ORİJİNAL ARAŞTIRMA/ARTICLE

The expert evaluation of Chat GPT-4 responses to surgery-related anxiety in patients who undergo Total Hip Replacement surgery

Özlem ŞAHİN AKBOĞA[1], Murat KORKMAZ[1], İrem Nisa KORKMAZ[3], Müberra KULU[4]
DOI: https://doi.org/10.5281/zenodo.13354314

[1]Faculty of Health Sciences, Department of Surgical Diseases Nursing, Yozgat Bozok Üniversity, Yozgat/Türkiye

[2]Faculty of Medicine, Department of Orthopaedics, Yozgat Bozok Üniversity, Yozgat/Türkiye

[3]Faculty of Medicine, Basic Medical Sciences Anatomy Department, Izmir Bakırçay University, İzmir/Türkiye

[4] Faculty of Medicine, Bakirkoy Mental Health and Diseases Hospital, 19 Mayis University, Samsun/Türkiye

Corresponding Author:
Özlem ŞAHİN AKBOĞA
ozlemsahin.os17@gmail.com

Geliş Tarihi/Received Date: 14.05.2024

Kabul Tarihi/Accepted Date: 14.06.2024

Yayınlanma Tarihi/Publishing Date:30.06.2024

Abstract

Chat GPT (Generative Pre-trained Transformer)-4, one of the fourth generation of artificial intelligence technologies, is a language model that generates written text using deep learning techniques. Chat GPT-4 can be used for various natural language processing tasks such as language translation, text summarisation and question answering. It is predicted that Chat GPT-4 is used by surgical patients. Many issues regarding the use of Chat GPT-4 are of interest. The study aimed to examine the adequacy of Chat GPT-4-generated responses to reduce anxiety in patients undergoing hip replacement surgery. Firstly, a qualitative interview was conducted with 12 patients who underwent hip surgery to understand their anxiety about the surgery. As a result of the interview, 15 contents related to their concerns about the surgery were determined. The obtained contents were typed into the Chat GPT-4 search engine. The answers generated by Chat GPT-4 about the content were sent to 15 experts in the field. A five-point Likert-type scale was used for the adequacy of the responses in reducing anxiety. Most of the experts (80.1%) stated that the answers to the questions to be asked to the GPT-4 were sufficient to reduce the anxiety of the patient. The remaining part of the experts reported that the answers of GPT-4 were inadequate. According to the results of the study, patients undergoing hip surgery can reduce their anxiety about the surgery by using the Chat GPT-4 search engine. Chat GPT-4 is a model that can be trained and developed. We think that the problems related to inadequacies will be eliminated with the developed versions of the model.

Key Words: *Anxiety, Artificial İntelligence, Chat GPT-4, Total Hip Arthroplasty*

Total Kalça Protezi cerrahisi geçiren hastalarda ameliyat ilişkin kaygıya yönelik Chat GPT-4'ün yanıtlarının uzman değerlendirmesi

Öz

Yapay zekâ teknolojilerin dördüncü nesillerinden biri olan Chat GPT (Generative Pre-trained Transformer)-4, derin öğrenme tekniklerini kullanarak yazılı metin oluşturan bir dil modelidir. ChatGPT-4, dil çevirisi, metin özetleme ve soru yanıtlama gibi çeşitli doğal dil işleme görevleri için kullanılabilir. Cerrahi hastalar tarafından Chat GPT-4'ün kullanıldığı öngörülmektedir. Chat GPT-4'ün kullanımına ilişkin birçok konu merak konusudur. Çalışma, kalça protezi cerrahisi geçiren hastaların ameliyata ilişkin Chat GPT-4'ün ürettiği yanıtların kaygıyı azaltmadaki yeterliliğini incelemeyi amaçladı. İlk olarak, kalça cerrahisi geçiren 12 hasta ile ameliyat ilişkin kaygılarını anlamaya yönelik niteliksel görüşme gerçekleştirildi. Görüşme sonucu ameliyata ilişkin kaygılarına yönelik 15 içerik belirlendi. Elde edilen içerikler Chat GPT-4 arama motoruna yazıldı. Chat GPT-4'ün içerik ile ilgili oluşturduğu yanıtlar alanında uzman 15 kişiye gönderildi. Yanıtların kaygıyı azaltmada yeterliliğine ilişkin beşli likert tipte ölçek kullanıldı. Uzmanların büyük bir kısmı (%80,1) hastanın kaygısını azaltmak için GPT-4'e soracağı sorulara ilişkin yanıtların yeterli olduğunu belirtti. Geriye kalan kısmı ise GPT-4'ün yanıtları açısından yetersizliğini bildirdi. Araştırmanın sonucuna göre, kalça cerrahisi geçiren hastalar ameliyata ilişkin kaygılarını Chat GPT-4 arama motorunu kullanılarak azaltılabilir. Chat GPT-4 eğitilebilmeye ve geliştirilmeye açık bir modeldir. Modelin geliştirilen sürümleri ile yetersizliklere ilişkin sorunların ortadan kalkacağını düşünmekteyiz.

Anahtar sözcükler: Chat GPT-4, kaygı, Total Kalça Protezi, yapay zekâ

Introduction

ChatGPT (Open AI, San Francisco, CA, USA) is a language model based on Artificial Intelligence (AI) technology, released to the public on November 1, 2022 (Open AI, 2023). The fourth generation of this model, which uses deep learning techniques to generate human-like texts (Lipman and Distler, 2023), is Generative Pre-trained Transformer (GPT-4). This artificial intelligence supported chatbot has features such as writing poems and articles and solving coding problems. The program develops quick and in-depth understanding of complex topics and generates specific conversational responses to the exact question asked, by recalling previous responses (Grant and Mets, 2022). The most common uses of ChatGPT include chatting, auto-writing, content creating and understanding language. The quality of the generated text is very good and even difficult to distinguish from a text written by humans (Lipman and Distler, 2023). ChatGPT, which is still being improved, is suitable for use in many areas. We believe that ChatGPT will help in many areas of medicine, which have complex problems difficult to understand.

One of the most difficult and complex surgeries of orthopaedics is Total Hip Replacement (THR). Today, in parallel with the prolongation of life expectancy and increase in musculoskeletal problems (osteoarthritis, rheumatoid arthritis, traumas, etc.), the number of THR surgeries has been gradually increasing (OECD, 2015). THR is a surgical treatment method that aims to restore hip movements and reduce pain by reconstructing joint surfaces. In health diagnosis before and after THR, patients are often found to have anxiety (Büyükyılmaz ve Güven Özdemir, 2018). Anxiety symptoms were found in 27.9% (149 patients) of the patients who underwent THR, while depression symptoms were found in 33.6% (Duivenvoorden et al., 2013). It is quite difficult to understand the cause of anxiety. Individuals who will receive THR are often afraid of dying, becoming disabled, not being able to support their families, losing their potential to work, experiencing pain after surgery, not being able to wake up after anaesthesia and losing control under the influence of anaesthesia (Yaban and Karaöz, 2007). This fear creates a vicious circle between pain and immobility and prepares the basis for the development of some postoperative complications (Monticone et al., 2013). In a study including approximately 8.700.000 patients who received total hip arthroplasty, serious complications ranging from anaemia to mortality at higher and more significant rates were reported in patients with high postoperative anxiety levels (Zalikha et al., 2021). This anxiety experienced by patients lead them to different searches. In this age of technology, the internet has become an important source of information where individuals access a large part of medical information (Erdoğan and Hocaoğlu, 2020). The use of ChatGPT, which can create quality texts about a given topic and which can answer health-related questions of individuals, is gradually increasing. In the light of this information, the

aim of the present study is to examine the efficiency of responses generated by Chat GPT-4 in reducing surgery-related anxiety of patients who undergo THR.

Material and Method

Design

The present study is a descriptive, cross-sectional research.

Determination of the Population and Sample

This study was carried out in the orthopaedics clinic of a university hospital in the Central Anatolia region of Turkey between 2022 and 2023 population of the study consisted of patients who underwent THR in this clinic. First, in-depth interviews were conducted with the patients who had undergone THR regarding their concerns about the surgery. The interviews were continued until the same types of answers were received from the participants in order to reach saturation (Polit and Beck, 2008). Inclusion criteria in the qualitative interviews were having undergone THR, and declaring voluntary participation in the study both orally and in writing. Exclusion criteria were having a history of psychiatric disease and drug use, having undergone surgery on the other hip, and lack of communication. Individuals who did not complete the interview were excluded. Secondly, 15 expressions related to anxiety were obtained from the interviews and the expressions were written in Chat GPT-4 program. Adequacy of the responses generated by Chat GPT-4 program in reducing the anxiety of patients who underwent THR was presented for the views of 15 experts.

Data collection tools

A questionnaire form prepared by reviewing the literature was used in data collection (Yaban ve Karaöz, 2007, Büyükyılmaz and Aştı, 2013, Wood et al., 2016). "Semi-structured Interview Form" was used in the questionnaire. The semi-structured interview form was prepared since there are no suitable valid and reliable standard measurement instruments questioning the anxiety of patients who will undergo THR surgery. Seven questions were asked in this part which were "What are your preoperative fears?", "What are your fears about the surgery?", "What are your fears about the post-operative process?", "What are your fears about walking after the surgery?", "What are your fears about daily living activities after surgery?" "What are your fears about your professional life after surgery?" and "What are your fears about pre-operative and post-operative death?" The views of five experts other than the researchers (an orthopedist, a psychiatrist, a psychologist, an academic nurse who has completed her doctorate and studies about orthopaedics clinic and physiotherapist) were taken for content validity and suitability of the data collection instrument and the experts scored the instrument as applicable. Therapeutically communication techniques were used during face-to-

face interviews with the participants, the participants were encouraged and motivated to speak freely. While one of the researchers took notes, the other recorded the interview, extra questions were asked to the participants depending of the flow of the interview. The interviews lasted for about 30-40 minutes and expressions to get response from ChatGPT were determined after the interview. The expressions were approved by the authors in addition to two senior researchers. Literature support was sought in determining the expressions in order to increase the transferability of research findings.

Data collection method

The expressions created were written to the address https://chat.openai.com/chat to get a response from ChatCPT. A 5-Likert type (very adequate, partly adequate, neutral, partly inadequate, inadequate) was used to determine the adequacy of ChatCPT response in reducing anxiety in patients who undergo THR. The responses of ChatCPT-4 and the scale items for each question were sent to 15 experts (12 psychiatrists and 3 psychologists). The results taken from experts who evaluated the responses of ChatCPT-4 were recorded.

Evaluation of data

A statistical package (IBM SPSS Statistics 26 (IBM SPSS, Turkey) program was used for statistical analysis. Descriptive statistics, mean, standard deviation and frequency analysis were performed.

Ethical aspects of the research

Permission was obtained from the relevant hospital and the local ethics commission to conduct the qualitative part of the study (Ethics Commission date=20.06.2023 and no=04/19). Participants were informed about the content and duration of the study and their consent for participation was obtained.

Results

As a result of the interviews with the patients, 15 expressions were found regarding their anxiety about THR surgery. Table 1 shows these expressions and the rates of expert views. The rates of adequacy given by experts for the responses of ChatCPT-4 are as follows: very adequate (80.6%), partly adequate (50%), neutral (6%), partly inadequate (2%) and inadequate (11.3%). The experts stated that ChatCPT-4 response of the expression “not being able to walk or meet needs as before after THR” was not very adequate (33.3%) (Table 1). In psychotherapy, under normal conditions, the therapist tries to learn the underlying meaning of the patient’s feelings of anxiety such as being needy and incapable of being self-sufficient and tries to solve these. According to experts, this low rate of proficiency was due to the fact that ChatCPT-4 response pushed back emotions, thoughts and subconscious system of patients. Experts criticized ChatCPT-4 for giving general information on the

topic. Similarly, expressions 6, 7, 11 and 12 written in ChatCPT-4 were on somatic concerns of patients. For this reason, the experts stated that ChatCPT-4 response was partly adequate or they were neutral about the adequacy of the response. According to the views of another expert about the responses by ChatCPT-4, it was emphasized that there is additional information that should be added and this information will increase the trust in the physician and therefore may be effective in reducing concerns. For example, the experts stated that in Expression 1, sharing information about mortality rate would emphasize that physicians decided for surgery by considering this rate and risk factors and information could be added in Expression 8 on research findings about to what extent THR decreased post-operative pain.

Table 1. Expressions and rates of experts views

	Very adequate n(%)	Partly adequate n(%)	Neutral n(%)	Partly inadequate n(%)	Inadequate n(%)
1. I am afraid of dying in THR surgery.	10 (66.6)	4 (26.6)	0 (0)	0 (0)	1(6.6)
2. I am afraid of not being able to support my family after THR surgery.	8 (53.3)	5 (33.3)	1(6.6)	0 (0)	1(6.6)
3. I am afraid of not being able to wake up after THR surgery.	9 (60)	4 (26.6)	1(6.6)	0 (0)	1(6.6)
4. I am afraid of not being able to walk or meet all my needs as before after THR surgery.	5 (33.3)	8 (53.3)	0 (0)	1(6.6)	1(6.6)
5. I think that I will have a lot of pain after THR surgery.	9 (60)	2 (13.2)	3 (20)	0 (0)	1(6.6)
6. I am afraid of becoming disabled after THR surgery.	7 (46.6)	7 (46.6)	0 (0)	0 (0)	1(6.6)
7. I am afraid of not being able to return to professional life after THR surgery.	7 (46.6)	6 (40)	1(6.6)	0 (0)	1(6.6)
8. I am afraid of not being able to get rid of all my pain and problems after THR surgery.	8 (53.3)	5 (33.3)	0 (0)	1(6.6)	1(6.6)
9. I think that I will die because of bleeding or another problem during THR surgery.	8 (53.3)	5 (33.3)	1(6.6)	0 (0)	1(6.6)
10. I think that I will have pain during THR surgery.	8 (53.3)	4 (26.6)	1(6.6)	1(6.6)	1(6.6)
11. I think that I will die because of bleeding or another problem during THR surgery.	7 (46.6)	7 (46.6)	1(6.6)	0 (0)	1(6.6)

12. I am afraid that my wound site will be infected or another problem will occur and I will not be able to heal after THR surgery.	7 (46.6)	6 (40)	1(6.6)	0 (0)	1(6.6)
13. I am afraid that my prosthesis will be dislocated after surgery.	8 (53.3)	5 (33.3)	0 (0)	0 (0)	2 (13.2)
14. I am afraid that my sex life will be adversely affected after THR.	9 (60)	4 (26.6)	0 (0)	0 (0)	2 (13.2)
15. I am worried that I will not be able to perform my religious services (such as prayers, ablution, etc.)	11 (73.3)	3 (20)	0 (0)	0 (0)	1(6.6)
Total	**%80.6**	**%50**	**%6**	**%2**	**%11.3**

*THR: Total Hip Replacement

Discussion

The present study researched the adequacy of responses generated by Chat GPT-4 in reducing the surgery-related anxiety of patients who undergo THR. We think that ChatGPT can also be used in orthopaedics surgery, as in other fields of medicine. In addition to being valuable for orthopaedics physicians and nurses working in the clinic, ChatGPT can also be a guide for patients. ChatGPT has been reported to be used for documenting structured data such as patient training, vital signs and laboratory results in clinical practice, drug management, researching and language translation (Moons and Bulck, 2024). In the present study, the use of ChatGPT in clinical practice for reducing anxiety, which is experienced in THR surgery and which is difficult to understand, was found to be adequate (42.6% very adequate, 37.3% adequate). It is a very difficult process to diagnose stress, which is an abstract concept, in patients who will undergo THR. Orthopaedics nurses and physicians have difficulties in managing the stress process (Sellevold et al., 2023). For this reason, it will be inevitable for healthcare providers to receive support from ChatGPT in managing patients' stress. This model, which was previously used in the triage field of health services (Baker et al., 2020), is open to development with studies conducted.

It can be said that ChatGPT is the strongest of these new generation chatbots supported by major language models (Stokel-Walker, 2022). It is likely that there will be concerns and reactions regarding the validity and reliability of technology despite major developments. In the present study, while there are expert views regarding the adequacy of ChatGPT response, recommendations regarding its inadequacy were also reported. ChatGPT response was found to push off the individual's emotions, thoughts and subconscious system or give general information regarding somatic concerns. We believe that for ChatGPT-4 to generate better quality response, information should be deepened and it should be taught the ability to ask questions regarding the subconscious of individuals. The use of computer technologies raises concerns about depersonalization in knowing the patient (Patel and Lam, 2023). The use of artificial intelligence has even been associated with outcomes that

compromise patient safety in previously conducted studies (Powles and Hodson, 2017). However, despite these negativities, we should continue to research the use of chatbots in the field of health services. For this reason, gathering expert views about the use of ChatGPT is essential in identifying deficiencies, arranging new generation versions, and addressing the concerns of healthcare professionals who will use ChatGPT. Concerns about the model should be understood and many obstacles should be overcome so that ChatGPT is placed in clinical practice.

Limitations of the Study

The present study has two limitations. The first one is that the expressions generated are limited to the views of patients. Patient expressions written to ChatCPT-4 are very superficial and general. The other limitation depending on this limitation is that the responses of ChatCPT-4 consist of written expressions.

Conclusion

According to the results of the present study, it can be said that the use of ChatGPT-4 is adequate to reduce the surgery-related anxiety of patients who undergo THR surgery. As in all societies, there are concerns about the use of ChatCPT-4 in our country, although they are low. The responses of ChatCPT-4 are spot information independent of the individual, without depth and without knowing the thinking system of the other person. To address these concerns, ChatCPT-4 should be developed further, it should be trained and gain the trust of healthcare providers. Deficiencies can be dealt with by publishing more studies. The introduction of ChatCPT-4 and its new generation versions into clinical practice will help healthcare professionals and the system in increasing the quality of patient care and patient satisfaction, and in reducing post-operative complications and therefore the cost of hospital. In addition to evaluating and training patients before THR surgery, answering patient questions at any time can reduce anxiety significantly. For this reason, patients should be able to access ChatCPT-4 whenever they want. In addition, the development and training of the deficiencies of ChatCPT-4 will expand its use in the other fields of medicine.

Acknowledgement

We would like to thank our experts who gave their suggestions in the article and our patients who underwent THR surgery and gave their opinions.

Conflict of interest

There is no conflict of interest between the authors.

Author contributions

OSA; data collection, data analysis, writing the article, final reading: MK; literature review, preparation of data collection forms, writing the findings, writing the article, final reading: INK; writing, writing the article, final reading: MK; data collection, final reading.

References

Baker, A., Perov, Y., Middleton, K., Baxter, J., Mullarkey, D., Sangar, D., ... & Johri, S. (2020). A comparison of artificial intelligence and human doctors for the purpose of triage and diagnosis. *Frontiers in Artificial İntelligence,* 3, 543405.

Büyükyılmaz, F., & Özdemir, N. G. (2018). New life with total hip and knee replacement: Key concepts in patient education. *Journal of Inonu University Vocational School of Health Services,* 6(2), 86-96. (in Turkish)

Büyükyılmaz, F., & Aştı, T. (2013). The effect of relaxation techniques and back massage on pain and anxiety in Turkish total hip or knee arthroplasty patients. *Pain Management Nursing,* 14(3), 143-154.

Duivenvoorden, T., Vissers, M. M., Verhaar, J. A. N., Busschbach, J. J. V., Gosens, T., Bloem, R. M., ... & Reijman, M. (2013). Anxiety and depressive symptoms before and after total hip and knee arthroplasty: a prospective multicentre study. *Osteoarthritis and Cartilage,* 21(12), 1834-1840.

Erdoğan, A., & Hocaoglu, C. (2020). Cyberchondria: A review. Current Approaches in Psychiatry, 12(4), 435-443. (in Turkish)

Grant, N., Mets, C. (2022). A new chat bot is a code red for Google's search business. The New York Times, Dec 21, 2022. https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html (accessed Feb 1, 2023).

Lipman, J., & Distler, R. (2023). Schools shouldn't ban access to ChatGPT: https://time.com/6246574/schools-shouldnt-ban-access-to-chatgpt/ (accessed Feb 3, 2023).

Monticone, M., Ferrante, S., Rocca, B., Salvaderi, S., Fiorentini, R., Restelli, M. and Foti, C. (2013). Home-based functional exercises aimed at managing kinesiophobia contribute to improving disability and quality of life of patients undergoing total knee arthroplasty: a randomized controlled trial. *Archives of Physical Medicine and Rehabilitation,* 94(2), 231-239

Moons, P., & Van Bulck, L. (2024). Using ChatGPT and Google Bard to improve the readability of written patient information: a proof of concept. *European Journal of Cardiovascular Nursing,* 23(2), 122-126.

OECD. (2015). Hip and knee replacement. In Health Care Activities Health at a Glance 2015: OECD Indicators. OECD Publishing, Paris, 112-113.

Open AI. (2022). ChatGPT: Optimizing language models for dialogue. OpenAI. https://openai.com/blog/chatgpt/ (accessed Jun 10, 2023).

Patel, S. B., & Lam, K. (2023). ChatGPT: the future of discharge summaries?. *The Lancet Digital Health,* 5(3), e107-e108.

Polit, D.F., & Beck, C.T. Nursing research: generating and assessing evidence for nursing practice. Lippincott Williams & Wilkins. 2008.

Powles, J., & Hodson, H. (2017). Google DeepMind and healthcare in an age of algorithms. *Health and Technology,* 7(4), 351-367.

Sellevold, V. B., Olsen, U., Lindberg, M. F., Steindal, S. A., Aamodt, A., Lerdal, A., & Dihle, A. (2023). "I am accustomed to something in my body causing pain": a qualitative study of knee replacement non-improvers' stories of previous painful and stressful experiences. *BMC Musculoskeletal Disorders,* 24(1), 1-11.

Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays-should academics worry?. *Nature*.

Yaban, Z. Ş., & Karaöz, S. (2007). Nursing care in total hip replacement surgery. *Journal of CÜ School of Nursing,* 11(1), 47-53.

Zalikha, A. K., Karabon, P., Hussein, I. H., & El-Othmani, M. M. (2021). Anxiety and depression impact on inhospital complications and outcomes after total knee and hip arthroplasty: a propensity score-weighted retrospective analysis. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons*, 29(20), 873-884.

Wood, T. J., Thornley, P., Petruccelli, D., Kabali, C., Winemaker, M., & de Beer, J. (2016). Preoperative predictors of pain catastrophizing, anxiety, and depression in patients undergoing total joint arthroplasty. *The Journal of Arthroplasty*, 31(12), 2750-2756.