



ORIGINAL ARTICLE

Medicine Science 2024;13(3):537-40

How much does ChatGPT know about breast cancer?

Sendag Yaslikaya, Mehmet Mutlu Kidi

Çukurova University, Faculty of Medicine, Medical Oncology Clinic, Adana, Türkiye

Received 04 January 2024; Accepted 24 June 2024

Available online 30 July 2024 with doi: 10.5455/medscience.2024.01.01

Content of this journal is licensed under a Creative Commons Attribution-NonCommercial-NonDerivatives 4.0 International License.



Abstract

As ChatGPT gained popularity, it attracted the attention of everyone in the medical field, including both medical professionals and patients who want to access information about diseases and treatment. Both patients and healthy people started to seek the shortest and fastest answers to their health-related questions with chatGPT. In this time when access to fast information has become easier, it has become difficult to distinguish correct information from incorrect information. In order not to get lost in this information confusion, it is necessary to be sure of the sources used. In our study, we wanted to examine the level of knowledge and adequacy of ChatGPT under headings such as general information about breast cancer, treatment options, risks, and ways of protection. We directed the 32 most frequently asked questions about breast cancer to chatGPT. The questions were asked in two languages, both Turkish and English. The answers received were evaluated by 2 experts. The answers given to the questions were generally considered sufficient. We evaluated it as a reliable source that can be used as it provides accurate and sufficient information to patients. However, we observed that using the English language rather than Turkish makes ChatGPT easier to understand and provides deeper and more accurate information.

Keywords: ChatGPT, breast cancer, artificial intelligence

Introduction

Chat GPT; is an artificial intelligence-based chatbot developed by OpenAI. Its name is GPT-3.5, which stands for "Generative Pre-trained Transformer 3.5". Its primary function is to understand and respond to text-based questions, generate text, or perform various language-related tasks. In fact, what Chat-GPT does is no different from other search engines. The robot can answer the questions you ask as if you were a human. Chat GPT understands which language you ask questions in, and gives you answers in that language. Its clear and fluent writing and the fact that it can be spoken in almost every language have made Chat-GPT usable in every subject. Chat GPT compiles the information obtained over the internet and presents it in a more summarized form. Chat GPT started to be used by people in many functional activities such as writing articles, creating projects, producing content, and learning languages. Not only these, they used it to find answers to all the questions in their mind. They provided access to a lot of information quickly and easily. However, it also raised concerns about the accuracy of Chat GPT. Evaluations

have shown that it made mistakes not only in article writing but also in mathematical calculations.

We asked Chat GPT the most frequently asked questions about breast cancer, which is the most common cancer in women worldwide and which not only patients but also healthy people may be curious about. We aimed to evaluate the accuracy of the answers we received in the presence of experts in the field and to guide recommendations.

Material and Methods

We asked ChatGPT the 32 most frequently asked questions about breast cancer. The questions were prepared based on the breast cancer patient guidelines of the European Society for Medical Oncology (ESMO) and National Comprehensive Cancer Network (NCCN) associations. Headings containing questions in the 'What is breast cancer: Let us answer some of your questions' guide prepared by ESMO for breast cancer patients were prioritized. Again, the topics that were particularly emphasized in the guide were turned into questions. The questions were asked in English

CITATION

Yaslikaya S, Kidi M. How much does ChatGPT know about breast cancer?. Med Science. 2024;13(3):537-40.



Corresponding Author: Sendag Yaslikaya, Çukurova University, Faculty of Medicine, Medical Oncology Clinic, Adana, Türkiye
Email: drysendag@gmail.com

as follows: 'What is breast cancer?', 'What is the frequency of breast cancer?', 'What are the risk factors for breast cancer?', 'What are the causes of breast cancer?', 'How is breast cancer diagnosed?', 'What are the symptoms of breast cancer', 'How will breast cancer treatment be determined?', 'What are the subtypes of breast cancer?', 'What are the treatment options for breast cancer?', 'What are the treatment options for non-invasive (Stage 0) breast cancer (also called in situ carcinoma or DCIS)?', 'What are the treatment options for early invasive (Stage I-IIA) breast cancer?', 'What are the treatment options for locally-advanced (Stage IIB-III) breast cancer?', 'What are the treatment options for metastatic (Stage IV) breast cancer?', 'My breast cancer treatment is over, can I get pregnant?', 'Is menopause important in breast cancer patients?', 'I have breast cancer, should I have genetic testing?', 'I had a genetic test and my BReast CAncer gene (BRCA) mutation was detected. What should I do next?', 'Are there any complications related to the treatments I receive?', 'After my treatment is completed, how often and how should my follow-ups be?', 'If my first degree relatives have breast cancer, what is my risk of getting breast cancer?', 'Is breast cancer seen in men?', 'How often and with what methods should we check to prevent breast cancer?', 'I do not want the entire breast to be removed, is this possible?', 'Does breast ultrasonography (USG) and mammography cause breast cancer?', 'I have fibrocystic disease. Will it turn into breast cancer?', 'I have fibroadenoma breast disease, will it turn into breast cancer?', 'Are there any clinical trial options for breast cancer?', 'could this treatment effect my sex life? If so how and how for long?', 'can breast cancer kill?', 'Are immunotherapies used in breast cancer?', 'what can i do to lower my changes of developing breast cancer', 'Will breast reconstruction surgery be performed?'. The questions were systematically divided into different groups according to relevant topics: basic information, cancer prevention, diagnosis, surgical treatment and recovery, medical treatment, treatment-related complications and follow-up. In practice, each question was asked once. The answers received were noted. The accuracy of the answers was evaluated by experts in the field in accordance with the ESMO and NCCN patient guidelines, which were used as sources in creating the questions. Experts' evaluations were evaluated according to four different options. In this study, we used ChatGPT (as per GPT 3.5, January 2022, OpenAI). The questions were published on December 14, 2023. Questions were independently reviewed and rated for accuracy and adequacy by two experienced medical oncologists. Questions were evaluated out of 4 points. Completeness of the answers analyzed based on the guides (100% correct and complete) was evaluated as '4 point-correct answer'. According to the guidelines, nearly complete answers (>50% correct information) were evaluated as '3 point-partially correct answer', while insufficient answers (<50% information) were evaluated as '2 point- inadequate answer'. In case of any misinformation, it was noted as '1 point-incorrect answer'. Each question was asked in both English and Turkish. The questions were asked consecutively at the same time interval. Since ChatGPT may generate different answers when the same questions are asked repeatedly or asked at different times, the questions were asked once in both Turkish

and English and the answers were noted. Over the notes, both experts evaluated the same answers and gave their scores. The answers received in both languages were noted and scored by experts. Since ChatGPT may have different answers to the same repetitive questions, the questions were asked both in Turkish and English at once, and the answers given were noted. Experts made their scores based on the noted questions and answers.

The data consists of English and Turkish language scores rated by Expert-1 and Expert-2. The analysis was conducted in three main stages: basic descriptive statistics, comparative analysis, and correlation analysis. The scores given by both experts were analyzed using independent sample and paired samples t-test. Basic Descriptive Statistics; The mean, median, standard deviation, minimum, and maximum values of the ratings were calculated for both English and Turkish scores. This provided an overview of the central tendency and dispersion of the ratings. Comparative Analysis; The scores of English and Turkish languages were compared to identify any significant differences. Additionally, the ratings given by Expert-1 were compared with those given by Expert-2 to assess inter-rater reliability. Correlation Analysis; The correlations between the ratings given by the different experts for both English and Turkish questions were examined. This analysis aimed to determine the degree of agreement between the experts.

Results

A total of 32 questions about breast cancer were posed to ChatGPT in both Turkish and English. The questions were categorized as follows: basic information (31%, 10 questions), cancer prevention (12.5%, 4 questions), diagnosis (3%, 1 question), surgical and medical treatment (31%, 10 questions), treatment-related complications and follow-up (21.8%, 7 questions).

For the English questions evaluated by Expert-1, 24 out of 32 questions were deemed completely correct. Among the other answers given to the questions, 7 out of 32 questions were partially correct, and 1 out of 32 questions was considered inadequate. There were no incorrect answers. The average score of the answers given to the English questions by Expert-1 was 3.71. For the English questions evaluated by Expert-2, 16 out of 32 questions were deemed completely correct. Among the other answers given to the questions, 16 out of 32 questions were partially correct. There were no inadequate or incorrect answers. The average score of the answers given to the English questions by Expert-2 was 3.46.

For the Turkish questions evaluated by Expert-1, 15 out of 32 questions were deemed completely correct. Among the other answers given to the questions, 14 out of 32 questions were partially correct, and 3 out of 32 questions were considered inadequate. There were no incorrect answers. The average score of the answers given to the Turkish questions by Expert-1 was 3.37. For the Turkish questions evaluated by Expert-2, only 2 out of 32 questions were deemed completely correct. Among the other answers given to the questions, 16 out of 32 questions were partially correct, and 14 out of 32 questions were considered inadequate. There were no incorrect answers. The average

score of the answers given to the Turkish questions by Expert-2 was 3.1. The score evaluation obtained from both Turkish and English responses of Expert 1 and Expert 2 is given in Figure 1. The results are summarized in Table 1.

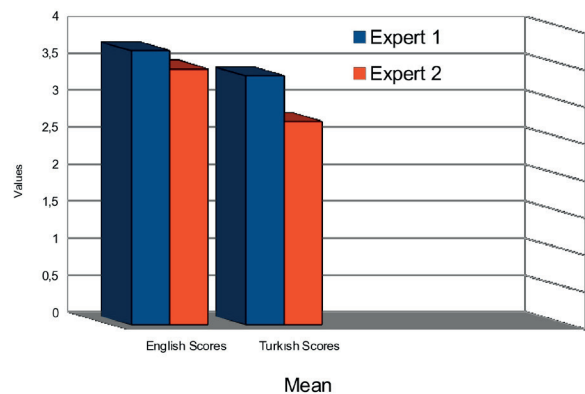


Figure 1. Experts Turkish and English scores

Table 1. Distribution of the responses from ChatGPT

	English		Turkish	
	Expert-1	Expert-2	Expert-1	Expert-2
Basic knowledge (n: 10);				
Correct	9	5	7	1
Partially correct	1	5	2	5
Mixed and misleading	0	0	1	4
Completely inaccurate	0	0	0	0
Cancer prevention (n: 4);				
Correct	4	3	3	2
Partially correct	0	1	1	2
Mixed and misleading	0	0	0	0
Completely inaccurate	0	0	0	0
Diagnosis (n: 1);				
Correct	1	1	1	1
Partially correct	0	0	0	0
Mixed and misleading	0	0	0	0
Completely inaccurate	0	0	0	0
Surgical/medical Treatment (n: 10);				
Correct	6	3	1	1
Partially correct	3	7	7	5
Mixed and misleading	1	0	2	4
Completely inaccurate	0	0	0	0
Treatment associated complications and follow-up (n: 7);				
Correct	1	2	1	0
Partially correct	1	0	1	2
Mixed and misleading	0	0	0	0
Completely inaccurate	0	0	0	0
Average score of the answers	3.37	3.16	3.71	3.46

The scores given by the two experts for the English and the Turkish questions were compared using a t-test. It was found that there was no statistically significant difference between the average scores given by Expert-1 and Expert-2 for the English questions. Similarly, it was determined that there was no statistically significant difference between the average scores given by Expert-1 and Expert-2 to the Turkish questions. When the scores given by the experts to both Turkish and English responses of ChatGPT were statistically evaluated, it was determined that there is a statistically significant difference in the response scores given by both Expert-1 and Expert-2 in both languages. When the correlation analysis of the scores given by the experts to both English and Turkish questions was evaluated, it was found that there is a statistically significant positive correlation between the responses of Expert-1 and Expert-2 in both Turkish and English.

Discussion

ChatGPT has become frequently used because it can answer questions in every field. In a time when it is so easy to access information, it becomes even more difficult to obtain accurate information. In our study, we evaluated the adequacy and accuracy of Chat GPT's answers to the most frequently asked questions about breast cancer. In our study, we directed the 32 most frequently asked questions about breast cancer to ChatGPT. We prepared our questions using the ESMO breast cancer patient guide and the NCCN breast cancer patient guides. Both the ESMO patient information guide and the NCCN patient guide have been prepared to help healthy people, those with relatives with breast cancer, and patients diagnosed with breast cancer who are trying to better understand its treatment. They contain information about early and late-stage breast cancer, including what causes the disease and how it is diagnosed, up-to-date guidance on the types of treatment that may be available and the possible side effects of treatment. The medical information described in these documents is based on the breast cancer Clinical Practice Guidelines, which are designed to assist physicians in the diagnosis and treatment of early and advanced breast cancer. The guides have been prepared both in the form of the most frequently asked questions and their answers and as a detailed information text, in a way that non-health individuals can easily understand [1,2]. We also scored the accuracy rates of our questions in light of these guides. When evaluated in the light of the guidelines, the answers provided by ChatGPT contained sufficient and accurate information. We think it is a reliable source to be used for information purposes. It has shown that it provides sufficient information for both healthy people and people receiving treatment. We showed that it provides more accurate and detailed information, especially in English compared to Turkish.

Health professionals have begun to use frequently artificial intelligence and machine learning in radiological identification, and histological evaluation. ChatGPT's knowledge and skills in the field of medicine have been evaluated by many different studies. For example, in a series of 100 cases where urological

cases were presented, it was shown that ChatGPTs responded correctly to 52% of the cases and correctly managed 11% of urgent urological cases [3]. In another study, treatment decisions of 10 breast cancer cases were compared with multidisciplinary tumor council decisions and ChatGPT answers. According to the study, although ChatGPT answers were generally responsive, accuracy rates were high. However, it turned out that he misjudged the low HER-2 positivity in patients [4]. In another study with head and neck cancer patients, ChatGPT answered 86.4% of the questions completely correctly, while there were no questions answered completely incorrectly [5]. Contrary to these mentioned studies, the ChatGPT study on hepatocellular cancer was shown to not provide sufficient information [6]. In our study, ChatGPT's answers to the questions were received full marks by both experts in at least half of them. We also think that ChatGPT is a good source for accessing detailed and sufficient information.

ChatGPT is used in translation from different languages: in a Chinese study testing whether ChatGPT grasps the language with questions shows that the model understands statistical concepts well, with a 93% correct answer rate [7]. The adequacy of ChatGPT's translation function was evaluated in another study, it was emphasized that although ChatGPT generally provides accurate translations, it needs to be adapted for some complex texts such as legal and medical texts [8]. However, in our study, ChatGPT's perception of questions in different languages differed. While the average score of our questions was determined to be higher in the answers given in English, the average score of the answers given in Turkish was found to be lower. While ChatGPT answers the same content questions in more detail in English, unfortunately, it is inadequate in Turkish. Therefore, we recommend that users use ChatGPT with caution in such studies. As in our study, ChatGPT failed to understand language differences arising from linguistic reasons in translation. When existing studies are evaluated, although there are conflicting results, it has been observed that ChatGPT is below the expected performance and should be used with caution. Accessing and directing accurate information becomes important, especially when considering cancer patients' search for alternative treatments other than their medical treatments.

Our study has limitations. We evaluated the latest version of ChatGPT (Chat GPT 3.5), which is available for free; Better results can be achieved in paid versions or developed versions. However, there will be a limitation in accessing the paid version. The topics and wording of the questions can be expressed in different ways. We designed most of the question wordings

from the topics and questions in the ESMO breast cancer patient guide. For this reason, we think that the answers obtained may differ if people ask different questions or express the questions differently. Again, the questions were asked and evaluated once. We think that doing multiple questions and answers may change the quality of the answers obtained. Again, scores were subjective and varied depending on the expert.

Conclusion

In conclusion, ChatGPT can reliably provide accurate information about breast cancer surveillance, diagnosis, and treatment, but both patients and professionals need to be cautious in their use of ChatGPT.

Conflict of Interests

The authors declare that there is no conflict of interest in the study.

Financial Disclosure

The authors declare that they have received no financial support for the study.

Ethical Approval

No patient was intervened and no patient information was used. Ethics committee approval was not received. Additionally, the article was designed and written by ethical rules.

References

1. Breast cancer: a guide for patients. <https://www.esmo.org/for-patients/patient-guides/breast-cancer> access date 21.05.2024
2. Guidelines for patients. <https://www.nccn.org/patientresources/patient-resources/guidelines-for-patients> access date 21.05.2024
3. Cocci A, Pezzoli M, Lo Re M, Russo GI, et al. Quality of information and appropriateness of ChatGPT outputs for urology patients. *Prostate Cancer Prostatic Dis.* 2024;27:103-8.
4. Lukac S, Dayan D, Fink V, et al. Evaluating ChatGPT as an adjunct for the multidisciplinary tumor board decision-making in primary breast cancer cases. *Arch Gynecol Obstet.* 2023;308:1831-44.
5. Kuşcu O, Pamuk AE, Sütay Süslü N, Hosal S. Is ChatGPT accurate and reliable in answering questions regarding head and neck cancer?. *Front Oncol.* 2023;13:1256459.
6. Cao JJ, Kwon DH, Ghaziani TT, et al. Accuracy of information provided by chatGPT regarding liver cancer surveillance and diagnosis. *AJR Am J Roentgenol.* 2023;221:556-9.
7. AL-Qadri AH, Ahmed SAM. Assessing the ChatGPT accuracy through principles of statistics exam: a performance and implications. *Research Square*, 2023 Mar 09. doi: 10.21203/rs.3.rs-2673838/v1. [Epub ahead of print].
8. Faten Khoshafah. ChatGPT for Arabic-English translation: evaluating the accuracy. *Research Square*. 2023 April 13. doi: 10.21203/rs.3.rs-2814154/v1. [Epub ahead of print].