

ChatGPT accurately performs genetic counseling for gynecologic cancers

Jharna M. Patel^{a,*}, Catherine E. Hermann^a, Whitfield B. Growdon^a, Emeline Aviki^b, Marina Stassenko^a

^a New York University Langone Health, Department of Obstetrics and Gynecology, Division of Gynecologic Oncology, New York, NY, United States of America

^b New York University Langone Health, Long Island, Department of Obstetrics and Gynecology, Division of Gynecologic Oncology, Mineola, NY, United States of America

HIGHLIGHTS

- ChatGPT provides comprehensive and correct responses to questions regarding genetic testing and counseling as it pertains to gynecologic oncology.
- ChatGPT can accurately provide genetic counseling, but further studies are necessary before it can be recommended as a patient resource.

ARTICLE INFO

Article history:

Received 22 January 2024

Received in revised form 3 April 2024

Accepted 8 April 2024

Available online xxxx

Keywords:

ChatGPT

Genetic counseling

Genetic testing

ABSTRACT

Objective. Artificial Intelligence (AI) systems such as ChatGPT can take medical examinations and counsel patients regarding medical diagnosis. We aim to quantify the accuracy of the ChatGPT V3.4 in answering commonly asked questions pertaining to genetic testing and counseling for gynecologic cancers.

Methods. Forty questions were formulated in conjunction with gynecologic oncologists and adapted from professional society guidelines and ChatGPT version 3.5 was queried, the version that is readily available to the public. The two categories of questions were genetic counseling guidelines and questions pertaining to specific genetic disorders. The answers were scored by two attending Gynecologic Oncologists according to the following scale: 1) correct and comprehensive, 2) correct but not comprehensive, 3) some correct, some incorrect, and 4) completely incorrect. Scoring discrepancies were resolved by additional third reviewer. The proportion of responses earning each score were calculated overall and within each question category.

Results. ChatGPT provided correct and comprehensive answers to 33/40 (82.5%) questions, correct but not comprehensive answers to 6/40 (15%) questions, partially incorrect answers to 1/40 (2.5%) questions, and completely incorrect answers to 0/40 (0%) questions. The genetic counseling category of questions had the highest proportion of answers that were both correct and comprehensive with ChatGPT answering all 20/20 questions with 100% accuracy and were comprehensive in responses. ChatGPT performed equally in the specific genetic disorders category, with 88.2% (15/17) and 66.6% (2/3) correct and comprehensive answers to questions pertaining to hereditary breast and ovarian cancer and Lynch syndrome questions respectively.

Conclusion. ChatGPT accurately answers questions about genetic syndromes, genetic testing, and counseling in majority of the studied questions. These data suggest this powerful tool can be utilized as a patient resource for genetic counseling questions, though more data input from gynecologic oncologists would be needed to educate patients on genetic syndromes.

© 2024 Elsevier Inc. All rights reserved.

1. Introduction

Artificial Intelligence (AI) is a growing field with significant applications in medicine. ChatGPT is a language AI model that uses an expansive data set and generates human-like responses. Its unique ability to

allow a user to steer a conversation has led to 180.5 million users as of November 2023 and over 1.7 billion queries each month [1]. OpenAI, the AI research group that founded ChatGPT, reports that there are ongoing limitations in its program and it has ongoing issues with perpetuating social bias as well as providing convincing but inaccurate responses because it does not provide source citations [1]. Despite this, ChatGPT is a growing source of information for patients and its utility in the medical field is an area of heightened interest [2]. A recent paper evaluating ChatGPT's ability to answer both fact based and counseling questions as it pertains to cervical cancer was recently

* Corresponding author at: NYU Langone Health, Department of OBGYN, Division of Gynecologic Oncology, 240 38th St, 20th Floor, New York, NY 10016, United States of America.

E-mail address: Jharna.Patel@nyulangone.org (J.M. Patel).

found to have mixed results. Though ChatGPT was able to accurately answer questions about cervical cancer prevention, survivorship, and quality of life, it was less accurate in its responses during the cervical cancer diagnosis and counseling questions [3].

We aim to further investigate ChatGPT's ability to answer counseling questions, specifically in genetic testing and counseling as it pertains to gynecologic oncology, as this field is very nuanced and crucial for personalized management based on individuals' genetic predispositions. Lynch syndrome and hereditary breast and ovarian cancer (HBOC) syndromes are the two most common genetic disorders encountered in gynecologic cancers. Lynch syndrome is a genetic condition that increases the risk of various cancers, particularly colorectal and endometrial cancers. HBOC syndromes involve mutations in the BRCA1 and BRCA2 genes, increasing susceptibility to breast and ovarian cancers, as well as other cancers. As genetic counseling is constantly evolving with concepts that are more challenging to patients, evaluating the accuracy of ChatGPT in answering commonly asked questions about genetic testing and counseling in relation to gynecologic cancers is important before it can be used as a potential tool for patients.

2. Methods

The study was exempt from the Institutional Review Board as it did not meet qualifications for Human Subjects Research. A list of 40 questions about genetic testing and counseling as it pertains to gynecologic oncology was formulated in conjunction with gynecologic oncologists and adapted from professional society websites. This included the American College of Obstetrics and Gynecology website, [CancerCare.org](#) and [Cancer.Net](#). Questions were formulated at an 8th-grade reading level in the first person to emulate an average patient asking questions as they would in a clinical setting. This would directly ChatGPT's ability to provide personalized answers to very specific questions regarding testing, result interpretation, and recommended surgical management. For example, one of the questions was "Should I have a hysterectomy if I have Lynch syndrome?" which would assess ChatGPT's ability to counsel a patient on the nuances of recommending a prophylactic hysterectomy if they have tested positive Lynch syndrome.

The questions were divided into two categories: general genetic counseling questions ($n = 20$) and syndrome specific questions about Lynch syndrome ($n = 3$) and HBOC syndrome ($n = 17$). Each question was queried using ChatGPT version 3.5, and each query was done in

privacy mode to eliminate biased answers from earlier queries. The answers were scored by two attending Gynecologic Oncologists (EA and MS) according to the following scale, adapted from previously published literature: 1) comprehensive and correct; there is no inaccurate information, and an attending Gynecologic Oncologist would likely have nothing significant to add, 2) correct, but not comprehensive; while there is no inaccurate information, an attending Gynecologic Oncologist would almost certainly have something to add, 3) some correct, some incorrect; contains some misinformation, 4) completely incorrect. A comprehensive list of questions can be found in [Table 1](#).

Scores were then evaluated for concordance and discrepant scores assigned to a question were sent to a third reviewer (WG) to assign a final score. If a final score could not be assigned due to different scores assigned by all reviewers, the question was excluded from final analysis. We calculated the proportion of ChatGPT responses that were given each score overall and within each question category. We also calculated the proportion of questions that required reviewers to resolve scoring discrepancies. A pre-planned secondary analysis was conducted in each question category and involved excluding questions that required a third reviewer to assign a final score.

3. Results

Out of 40 questions, half ($n = 20$) of the questions were categorized under general genetic counseling and the other half ($n = 20$) of questions pertained to Lynch and HBOC syndrome. Examples of queries that were assigned a final score of 1, 2, and 3 are listed in [Table 2](#). Overall, ChatGPT provided correct and comprehensive answers to 82.5% ($n = 33/40$) of questions, correct but not comprehensive answers to 15% ($n = 6/40$) of questions, partially incorrect answers to 2.5% ($n = 1/40$) of questions, and provided 0% ($n = 0/40$) incorrect answers. There were no questions that required more than three reviewers to assign a final score. Concordance rates between reviewers overall was 92.5% ($n = 37/40$).

Of the general genetic counseling questions, ChatGPT provided correct and comprehensive answers to 100% ($n = 20$) of questions with a 100% concordance rate between the two reviewers.

In the subset of questions that pertain to a specific syndrome (Lynch or HBOC), 17 questions were about HBOC syndrome and 3 questions about Lynch syndrome. Of the HBOC syndrome questions, ChatGPT provided correct and comprehensive answers to 64.7% ($n = 11/17$) of

Table 1
List of questions used to query ChatGPT.

Questions pertaining to general genetic counseling	Questions pertaining to genetic testing of specific syndromes	
	BRCA testing	Lynch Syndrome
What are genetic disorders?	What is hereditary breast and ovarian cancer syndrome?	How does lynch syndrome increase your risk for endometrial cancer?
What are genes?	What are BRCA1 and BRCA2?	If I have endometrial cancer do I also have lynch syndrome?
What are chromosomes?	How common are BRCA mutations?	Should I have a hysterectomy if I have lynch syndrome?
How do you know if someone has inherited or family cancer syndrome?	How much do BRCA mutations increase the risk for breast cancer?	
What are family cancer syndromes?	How much do BRCA mutations increase risk for ovarian cancer?	
What is genetic testing for cancer?	Do BRCA mutations increase risk for other types of cancer?	
Who needs to get genetic testing?	Who should be tested for BRCA mutations?	
How is genetic testing performed?	How is testing for BRCA mutations done?	
What are the types of genetic tests?	What does a positive test result for BRCA mutations mean?	
What should I know before getting genetic testing?	What does a negative test result for BRCA mutations mean?	
What is genetic counseling?	What does an indeterminate test result for BRCA mutations mean?	
How is genetic testing completed?	What screening tests are available for people with BRCA mutations?	
What is multigene panel testing?	How can you prevent cancer if you test positive for BRCA mutations?	
Who needs to get multigene panel testing?	Does having your ovaries surgically removed completely reduce the risk of ovarian cancer in women with BRCA mutations?	
How will genetic testing results affect my family?	Do BRCA mutations increase your risk for uterine cancer?	
Will genetic testing lead to more tests?	Should women with BRCA1 mutations have a hysterectomy?	
Who pays for genetic testing?	Should women with BRCA2 mutations have a hysterectomy?	
Could genetic tests lead to discrimination?		
What if genetic testing shows an increased cancer risk?		
Should I share results with my family members?		

Table 2
Examples of responses given by ChatGPT.

Question	Response Provided by ChatGPT	Final Score Assigned
Should I share genetic testing results with my family members?	<p>Cascade Testing: Cascade testing involves offering genetic testing to at-risk family members based on the identified genetic variant. By sharing your genetic testing results, you enable your family members to access testing that can provide them with important health information. This can lead to earlier detection, enhanced surveillance, or preventive interventions, which may improve health outcomes.</p> <p>Informed Decision Making: Genetic testing results can inform family members' healthcare decisions, including preventive measures, screening schedules, and treatment options. Sharing your results allows them to make informed choices about their healthcare and potentially benefit from personalized risk management strategies.</p> <p>Emotional Support: Genetic testing results may have emotional and psychological implications for individuals and their family members. By sharing your results, you provide an opportunity for mutual support and understanding within the family. Sharing experiences, concerns, and coping strategies can help alleviate anxiety and foster a supportive environment.</p> <p>When considering sharing genetic testing results with family members, it's important to do so in a sensitive and thoughtful manner. Here are a few suggestions:</p> <p>Choose the Right Time and Setting: Find an appropriate time and place to discuss the results with your family members. Consider their emotional readiness and ensure privacy for open and honest conversations.</p> <p>Provide Information and Resources: Explain the implications of the genetic testing results in a clear and understandable way. Offer educational materials or direct them to reliable sources of information to help them comprehend the significance of the findings.</p> <p>Respect Individual Choices: Understand that each family member has the right to make their own decisions about genetic testing or subsequent healthcare actions. Respect their autonomy and provide support without exerting pressure.</p> <p>Genetic Counseling: Encourage family members to seek genetic counseling or consult with healthcare professionals experienced in genetics. These professionals can help interpret the implications of the genetic testing results and guide family members through their own decision-making processes.</p> <p>Ultimately, the decision to share genetic testing results with family members is a personal one. It's important to weigh the potential benefits and consider the dynamics of your specific family situation. Genetic counseling can be instrumental in navigating these conversations and supporting you in making informed decisions about sharing your genetic testing results.</p>	1
Should women with BRCA1 mutations have a hysterectomy?	<p>The decision to undergo a hysterectomy (removal of the uterus) for women with BRCA1 mutations is a complex and individualized one that should be made in consultation with healthcare providers, genetic counselors, and based on a thorough assessment of personal risk factors. It is not a standard recommendation for all women with BRCA1 mutations.</p> <p>BRCA1 mutations primarily increase the risk of breast and ovarian cancers, rather than uterine (endometrial) cancer. The risk of developing uterine cancer in women with BRCA1 mutations is generally considered to be low or not significantly elevated compared to the general population. Therefore, a hysterectomy is not routinely recommended solely for the purpose of reducing uterine cancer risk in women with BRCA1 mutations.</p> <p>However, there may be other factors or individual circumstances that could influence the decision to undergo a hysterectomy. For instance, if a woman has other risk factors for uterine cancer, such as a personal or family history of the disease, or if she has other medical conditions that necessitate a hysterectomy, the procedure may be considered. It is crucial for women with BRCA1 mutations to discuss their specific risks, concerns, and options with their healthcare providers. They may also benefit from genetic counseling to better understand their individual risk profiles and make informed decisions about cancer risk reduction strategies, including the potential role of a hysterectomy in their particular case.</p>	2
What screening tests are available for people with BRCA mutations?	<p>For individuals with BRCA mutations, regular and specialized screening tests are available to detect and monitor for early signs of breast, ovarian, and other associated cancers. The specific screening recommendations may vary based on individual circumstances, medical history, and the recommendations of healthcare professionals. Here are some common screening tests for people with BRCA mutations:</p> <p>Breast Cancer Screening:</p> <p>Mammography: Regular mammograms are recommended for women with BRCA mutations, typically starting at an earlier age (around 25–30) than the general population. Annual or biennial mammograms may be recommended, depending on the individual's risk profile.</p> <p>Breast MRI: Magnetic resonance imaging (MRI) is often recommended in addition to mammography for breast cancer screening in individuals with mutations. Breast MRI is more sensitive than mammography and can help detect breast tumors that may be missed by mammograms, particularly in younger women with denser breast tissue. Screening intervals may vary, but annual breast MRI is commonly recommended.</p> <p>Clinical Breast Exams (CBE): Regular clinical breast exams by a healthcare professional are typically recommended every 6 to 12 months. CBE involves a physical examination of the breasts and surrounding areas to check for any abnormalities.</p> <p>Ovarian and Fallopian Tube Cancer Screening:</p> <p>Transvaginal Ultrasound (TVUS): Transvaginal ultrasound involves the use of an ultrasound probe inserted into the vagina to obtain images of the ovaries and fallopian tubes. It may be used as part of ovarian cancer screening for individuals with BRCA mutations, although its effectiveness in detecting early-stage ovarian cancer is limited.</p> <p>CA-125 Blood Test: CA-125 is a tumor marker that can be measured through a blood test. Elevated levels of CA-125 may indicate the presence of ovarian cancer. However, CA-125 levels can be influenced by various factors, and the test is not specific enough to detect early-stage ovarian cancer reliably. Therefore, it is not recommended as a standalone screening test but may be used in conjunction with other screening methods.</p> <p>It's important to note that while these screening tests can aid in the detection of cancer at an earlier stage, they do not eliminate the risk of developing cancer. Risk-reducing strategies, such as prophylactic surgeries (mastectomy, salpingo-oophorectomy), may be discussed as additional options to further reduce cancer risk. Regular consultations with healthcare professionals and genetic counselors are crucial to determining the most appropriate and personalized screening strategies based on an individual's risk profile and preferences.</p>	3

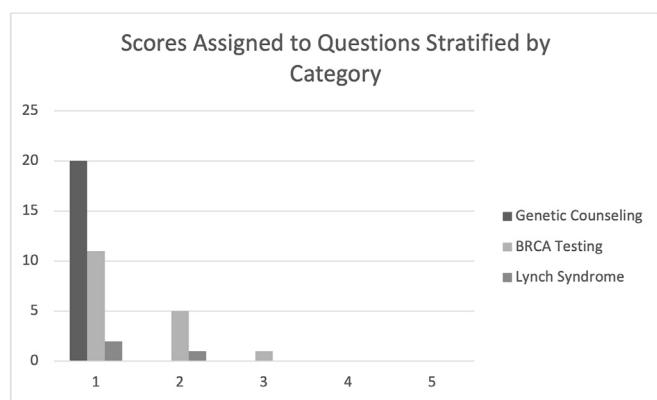


Fig. 1. Scores assigned to questions, stratified by category. Table 2 examples of responses given by ChatGPT.

questions, correct but not comprehensive answers to 29.4% ($n = 5/17$) of questions, and partially incorrect answers to 5.8% ($n = 1/17$) of questions. No completely incorrect answers were reported. Concordance rates between the two reviewers reached 88.2% ($n = 15/17$) with only 2 questions requiring a third reviewer for scoring.

In questions about Lynch syndrome, ChatGPT provided correct and comprehensive answers to 66.6% ($n = 2/3$) of questions and correct but not comprehensive answers to 33.3% ($n = 1/3$) of questions. Concordance rates between the two reviewers reached 66% with only 1 question sent to the third reviewer for scoring. The scoring breakdown for each category is provided in Fig. 1.

In secondary analysis, 3 questions from the syndrome specific question category that required a third reviewer were excluded to evaluate inter-reviewer reliability. This analysis found that ChatGPT provided correct and comprehensive answers to 70.6% ($n = 12/17$), correct but not comprehensive answers to 23.5% ($n = 4/17$) of questions, and partial incorrect answers to 5.9% ($n = 1/17$) of questions. This was similar to our initial analysis, highlighting the low rate of discrepancy between two primary reviewers.

4. Discussion

Our study indicates that ChatGPT can accurately provide genetic counseling and answer questions regarding general genetic testing but is less accurate in questions pertaining to syndrome specific genetic testing.

To our knowledge, this is a novel study evaluating ChatGPT's ability to provide genetic counseling for patients who may be at increased risk of developing gynecologic cancers. Recently, there has been an influx of studies geared toward the evaluation of ChatGPT's role in medicine. Many studies have had varied outcomes in ChatGPT's ability to answer questions among different fields.

This group has also conducted research in assessing ChatGPT's ability to answer commonly asked questions about cervical cancer prevention, diagnosis, treatment, and survivorship. ChatGPT performed the best in the cervical cancer prevention and survivorship categories but was less accurate in providing answers regarding treatment [3]. Of the three questions were asked in the diagnosis category but ChatGPT only answered one of them correctly [3]. ChatGPT performed best in questions that were strictly factual and was less accurate in answering nuanced questions that would be catered to a patient's situation [3].

In another study to assess ChatGPT's ability to provide knowledge and management of cirrhosis and hepatocellular carcinoma (HCC), ChatGPT regurgitated knowledge of cirrhosis and HCC, but only small proportions of responses were labeled as comprehensive [4]. When individual question categories were analyzed, ChatGPT performed better in knowledge, lifestyle, and treatment questions but was less accurate

in preventative questions, in contrast to the findings when evaluating its ability to answer questions about cervical cancer [3,4]. Both studies were conducted using similar methodology and scoring system which highlights the inconsistencies in ChatGPT's performance. In our study evaluating ChatGPT's ability to answer nuanced general genetic counseling and syndrome specific questions, 97.5% ($n = 39/40$) of questions were labeled as comprehensive and complete or correct but incomplete. These conflicting findings have made it difficult to define ChatGPT's applicability in medicine.

Still, several studies investing its applications in image analysis, screening patients for clinical trials, and aiding in medical diagnoses are underway [5]. Our findings indicate that ChatGPT is accurate and consistent in providing general counseling regarding genetic testing, indicating it can become a powerful tool to aid physicians and genetic counselors, especially as the model continues to improve.

A strength of our study is its quantitative analysis of ChatGPT's ability to provide genetic counseling. A full panel of questions was generated using professional sites with gynecologic oncologists who frequently field such questions in their clinical practice. The questions were both aimed at the financial implications of genetic testing (like insurance coverage), and questions that were pertinent to specific genetic syndromes. Each question was phrased at a reading level which would represent an average ChatGPT user. All reviewers (EA, WG, and MS) are board certified Gynecologic Oncologists from robust academic practices, lending credibility to the scoring system. Our study also had a low rate of inter-reviewer variability, with only 7.5% ($n = 3/40$) questions requiring additional reviewers to assign a final score.

This scoring system was derived from other studies of a similar objective, which could be a limitation of the study, as all scoring systems are subject to inherent scoring bias. Our study attempted to limit this by blinding each reviewer's responses. Another limitation of the study is the number of questions on Lynch syndrome ($n = 3$) as it pertains to gynecologic malignancies compared to HBOC syndrome ($n = 17$) questions which really limits any conclusions we can make about ChatGPT's ability to answer Lynch syndrome questions compared to HBOC syndrome questions. Initially, more questions about Lynch syndrome were incorporated but after review decision was made to only query those questions pertaining to gynecologic oncology. The addition of more questions about Lynch syndrome could allow us to compare these questions with HBOC syndrome questions directly.

The ChatGPT model 3.5 is based on 175 billion parameters compared to a subscription model 4.0 which has 1 trillion parameters and is plausible that the 3.5 model produces fewer instances of factual responses [1]. Our study utilized ChatGPT model 3.5 as it is most widely accessible to the average user. It should also be noted that ChatGPT does not provide resources as to where it obtained its responses which highlights that the program can provide convincing but inaccurate responses.

Future studies would need to evaluate the safety and accuracy of different ChatGPT models and how physician input could improve the accuracy of ChatGPT and improve its utility for patients. As our understanding of AI evolves, we must continue to evaluate how it can be applied to help improve patient care. We conclude that ChatGPT can accurately provide genetic counseling, but further studies are necessary before it can be recommended as a patient resource.

Funding

None.

CRediT authorship contribution statement

Jharna M. Patel: Writing – original draft, Project administration, Formal analysis, Data curation, Conceptualization. **Catherine E. Hermann:** Writing – review & editing, Formal analysis, Conceptualization. **Whitfield B. Growdon:** Data curation. **Emeline Aviki:** Writing – review & editing, Supervision, Data curation, Conceptualization.

Marina Stasenکو: Writing – review & editing, Supervision, Data curation, Conceptualization.

Declaration of competing interest

The authors have no conflicts of interest.

References

- [1] OpenAI ChatGPT: optimizing language models for dialogue. OpenAI, 2022 Nov, <https://openai.com/blog/chatgpt>.
- [2] T. Dave, S.A. Athaluri, S. Singh, ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations, *Front. Artif. Intell.* 6 (2023 May 4) 1169595, <https://doi.org/10.3389/frai.2023.1169595>. PMID: 37215063; PMCID: PMC10192861.
- [3] C.E. Hermann, J.M. Patel, L. Boyd, W.B. Growdon, E. Aviki, M. Stasenکو, Let's chat about cervical cancer: assessing the accuracy of ChatGPT responses to cervical cancer questions, *Gynecol. Oncol.* 179 (2023) 164–168.
- [4] Y.H. Yeo, J.S. Samaan, W.H. Ng, et al., Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma, *Clin. Mol. Hepatol.* 29 (3) (2023) 721–732, <https://doi.org/10.3350/cmh.2023.0089>.
- [5] A. Grunebaum, J. Chervenak, S.L. Pollet, A. Katz, et al., The exciting potential for ChatGPT in obstetrics and gynecology, *Am. J. Obstet. Gynecol.* 228 (6) (2023 Jun) 696–705.