ORIGINAL STUDY

Oman Medical Journal [2026], Vol. 41, No. 1: e795

Assessing the Accuracy of ChatGPT's Answers to Basic Questions on Uterine and Cervical Cancers

Aqdas Malik[1]*, Rashida Suleiman[2], Saria Bala[2], Shima Ibrahim[3], Moza Al Kalbani[2], Hilal Al-Busaidi[1] and Ikram A. Burney[2]

[1]Department of Information Systems, Sultan Qaboos University, Muscat, Oman
[2]Women's Health Program, Sultan Qaboos Comprehensive Cancer Care and Research Centre, Muscat, Oman
[3]Department of Gynecological Oncology, Bezmialem Vakıf University, Istanbul, Turkey

ARTICLE INFO

Article history:
Received: 13 March 2025
Accepted: 2 December 2025

Online:
DOI 10.5001/omj.2025.106

Keywords:
Gynecologic Neoplasms; Uterus; Cervix Uteri; Artificial Intelligence; Large Language Models; ChatGPT.

ABSTRACT

Objectives: Artificial intelligence (AI) platforms based on large language models such as ChatGPT, are increasingly being used by both the public and medical professionals to obtain medical information. This rapid growth in reliance makes it essential to systematically evaluate the accuracy and clinical reliability of AI-generated medical content. The objective of this study was to evaluate the accuracy of responses provided by ChatGPT regarding prevention, screening, treatment, and risk factors of common gynecological cancers. The assessment focused primarily on the use of ChatGPT by primary care providers and the public with limited subject-specific knowledge. ***Methods***: We evaluated the reliability of ChatGPT (version 3.5) in answering questions about two of the most common gynecological cancers. ChatGPT was posed a total of 40 questions on the prevention, screening, and treatment of endometrial cancer (20 questions) and cervical cancer (20 questions). Responses were independently reviewed and categorized as accurate, inadequate, or inaccurate by five physicians with a mean of 18 ± 3 years of experience in gynecological oncology. Reviewers provided reasons for deeming some responses as inadequate or inaccurate. ***Results:*** Overall, 20 out of 40 (50%) responses by ChatGPT 3.5 were regarded as either inaccurate or inadequate. Most of the deficient responses were related to questions on the treatment of the two cancers, while responses to questions about prevention were mostly accurate. ***Conclusions:*** ChatGPT may provide accurate information about prevention of gynecological cancers, but the public and health professionals should not rely on its responses to make medical decisions, as many responses in this domain were inadequate or inaccurate. Consultation with qualified physicians or specialists is essential for individualized decision-making. Medical information sourced from AI tools such as ChatGPT should be integrated with clinician oversight to improve reliability.

Gynecological cancers collectively constitute the fifth most common cancers worldwide.[1] Public awareness regarding their prevention, early detection/diagnosis, and treatment has increased substantially. High-quality information for patients and caregivers is widely available in print and digital media.[2] Authoritative bodies such as the World Health Organization publish evidence-based information on prevention, screening, and management of gynecological cancer. However, the information they provide is general and may not be applicable to individual cases.[3]

The unprecedented availability and accessibility of health information on the internet have significantly transformed how the public engages with cancer prevention and management. Online platforms such as YouTube, search engines such as Google, and health information websites like WebMD have become widely used sources of information on cancer prevention, screening, early detection, and treatment.[4] This easy access has enabled patients and the general public to obtain disease-related information remotely, improve their understanding of health conditions, develop self-management skills, and enhance survivorship awareness.[5] However, there are concerns about the accuracy and clinical reliability of internet-based medical information.

More recently, the public release of large language model (LLM)-based conversational systems has fundamentally transformed how medical information is accessed and consumed.[6] These systems have shown

*Corresponding author: ✉a.malik@squ.edu.om

strong potential for patient education, cancer screening, symptom monitoring, and survivorship care, ultimately improving quality of care.[6] ChatGPT is one of the leading conversational AI platforms, with hundreds of millions of active users. An increasing number of people now use ChatGPT to access information related to medical conditions, including cancer. In clinical scenarios, ChatGPT has been reported to perform better than or on par with other AI systems and even human physicians.[7] However, the model has also been reported to perform variably depending on the task assigned.

Perera Molligoda Arachchige,[7] studied ChatGPT's strengths and limitations in providing information related to emergency radiology and reported that it enhanced medical workflows through report generation, increased efficiency, and reduced clinician workload, despite occasional false positives. More studies have reported that ChatGPT showed promising capability in retrieving key oncological information.[8] In medical oncology settings, the platform's strengths include reviewing and summarizing patient records, interpreting next-generation sequencing reports, and suggesting clinical trial options. ChatGPT also demonstrated accuracy in providing information about common cancer myths and misconceptions that aligned closely with answers provided by the National Cancer Institute, though its treatment recommendations were not always consistent with National Comprehensive Cancer Network guidelines.[9,10] For example, the accuracy and reliability of responses related to the prevention, surveillance, and diagnosis of liver, prostate, and colon cancers were deemed suboptimal.[11–14]

In the current study, we report the accuracy and reliability of ChatGPT in providing information about prevention, early detection and screening, diagnosis, and stage-specific treatment choices for the two most common gynecological cancers: uterine cancer and cervical cancer. The primary objective is to assess how accurate the responses are when the public or primary care practitioners with limited subject knowledge consult the free version of ChatGPT.

METHODS

Two authors, both experienced in delivering continuing medical education in gynecological oncology to primary care practitioners, systematically reviewed and compiled the most frequent questions patients asked primary care providers regarding cervical and uterine cancer. The authors also used their own clinical experience to generate additional questions. Each author independently developed 40 questions (20 questions on endometrial cancer and 20 on cervical cancer). A third author reviewed and refined the questions. The finalized questions were submitted to ChatGPT version 3.5 on March 19, 2024.

ChatGPT's responses were independently assessed by five physicians from academic centers in two countries (one medical oncologist and four gynecological oncologists) using a standard three-tier classification:[15–17] (a) 'accurate' (all information is correct and relevant); (b) 'inadequate' (information is correct, but incomplete or irrelevant); or (c) 'inaccurate' (information is incorrect). Scores of +1, 0, and -1 were assigned, respectively. The mean score for each response was classified as accurate (mean score > 0.5), inadequate (-0.5–0.5), and inaccurate (< -0.5). The reviewers were asked to provide their reasons for 'inadequate' and 'inaccurate' ratings. This study did not require institutional ethical approval because no human subjects were involved.

RESULTS

The median age of the five reviewers was 45 years (range = 40–61); they had a mean experience of 18 ± 3 years in gynecological oncology. Table 1 shows questions, mean scores, and reviewers' explanations for inadequate or inaccurate responses by ChatGPT for cervical cancer. Table 2 shows the same information for uterine cancer.

For questions related to cervical cancer, two responses were rated 'inaccurate' (both related to treatment) and eight 'inadequate'. The overall reliability rate (accurate response = overall score > 0.5) was 50% (10/20). For questions related to endometrial cancer, 10 responses were regarded 'inadequate' (mainly related to treatment). Here also, the overall reliability rate was 50% (10/20).

The inter-coder reliability of the reviewing physicians' assessments of responses of ChatGPT was evaluated using Krippendorff's alpha, a robust statistical measure to evaluate ordinal data and multiple raters, which is also less sensitive to the number of raters or categories. Krippendorff's alpha was 0.85 for cervical cancer questions, indicating high agreement, and 0.78

Table 1: Cervical cancer related questions, mean reviewer scores, and comments on ChatGPT responses.

S. No.	Questions	Scores	Comments on inadequate or inaccurate responses
1.	What are the three most common symptoms of cervical cancer?	1	
2.	Is there anything I can do to prevent cervical cancer?	1	
3.	What is the relationship between human papilloma virus and cervical cancer?	1	
4.	How is cervical cancer treated?	0.4	Trachelectomy was not mentioned as an option for early stage and as a fertility-sparing technique. Radical hysterectomy was incorrectly stated as a treatment for advanced disease.
5.	Who is eligible for a cervical cancer screening?	0.8	Missing information: All individuals with a cervix are eligible for screening regardless of sexual activity, sexual orientation, or gender identity.
6.	What are the options for cervical cancer screening?	0.4	HPV testing was not specified as the gold standard. The fact that VILI and VIA are used only in low-resource-settings was not mentioned.
7.	Are the results of cervical screening reliable?	0.4	Sensitivity and specificity of the test were not mentioned.
8.	At what age should screening for cervical cancer start?	0.8	Conditions for screening regardless of sexual activity were not mentioned.
9.	How frequently should a person undergo cervical cancer screening?	1	
10.	Dose pregnancy increase the risk of developing cancer of cervix?	1	
11.	What is the effect of treatment of cervical cancer on subsequent pregnancy?	0.2	The possibility that chemoradiation may cause infertility was not mentioned.
12.	What would happen if the screening for cervical cancer showed an abnormal result?	0.4	Simple hysterectomy was not specified as an option for women who have completed childbearing.
13.	What are the different stages of premalignant lesion of the cervix?	1	
14.	What are the treatment options for premalignant lesion of the cervix?	0.2	Not specifically mentioned that hysterectomy for CIN3 is more suited for women who have completed childbearing.
15.	What are the treatment options of early-stage cervical cancer?	-0.8	The following was not made clear: Early-stage cervical cancer is typically classified as either stage IA (confined to the cervix) or stage IB (spread beyond the cervix but not to nearby organs). The treatment of choice is surgery. Chemoradiation is not often used for early-stage disease.
16.	What are the treatment options for locally advanced cervical cancer?	-0.6	Unclear or missing information: Locally advanced disease refers to stage IIB, Stage III, or stage IVA disease. Chemoradiation is the treatment of choice. Mention of surgery, targeted therapy, or clinical trial could be misinterpreted.
17.	What are the treatment options for metastatic and recurrent cervical cancer	0.2	Preferred options and indications for immune checkpoint inhibitor, such as CPS were not mentioned.
18.	What is the risk involved in undergoing surgery for cervical cancer	0.2	Not mentioned; DVT, ureteric, bladder and bowel injury.
19.	What are the side effects of radiotherapy for cervical cancer?	0.8	Not mentioned: specificity or severity of possible side effects.
20.	What are the side effect of chemotherapy for cervical cancer?	0.6	Risk of severe side effects (e.g., peripheral neuropathy, cutaneous toxicity) were not mentioned

HPV: human papillomavirus; VIA: visual inspection with acetic acid; VILI: visual inspection with Lugol's iodine; CIN: cervical intraepithelial neoplasia; CPS: combined positive score; DVT: deep vein thrombosis.

for uterine cancer questions, reflecting substantial agreement. These findings suggest that ratings of physicians were reliable and consistent across both sets of questions.

DISCUSSION

Overall, 20 (50%) ChatGPT responses were classified as either 'inaccurate' (factually incorrect) or 'inadequate' (correct, but incomplete or

Table 2: Uterine cancer related questions, mean reviewer scores, and comments on ChatGPT responses.

S. No.	Questions	Scores	Comments on inadequate or inaccurate responses
1.	What are the risk factors for uterine cancer?	0.4	Missing risk factors: early menarche and late menopause, genetic mutation (other than MMR gene), and medical conditions such as liver cirrhosis due to impaired estrogen metabolism.
2.	What common genetic mutations are associated with uterine cancer?	0.8	Not mentioned: Specific mutations associated with uterine cancer.
3.	Who should have genetic testing for uterine cancer?	0.6	Not mentioned: Early onset bilateral cancers are relevant for ovarian cancer, not uterine cancer.
4.	What is the standard screening test for endometrial cancer?	0.6	Not clarified: TVUS and biopsy are diagnostic tests, not screening tests for general population.
5.	What is the risk of developing endometrial cancer in patients who take tamoxifen for treatment of breast cancer?	0	Not mentioned: Estimated rate of risk and the fact that tamoxifen-associated endometrial cancer has poorer prognosis than endometroid type.
6.	What is the lifetime risk of endometrial cancer in people known to have Cowden syndrome?	0.6	The lifetime risk was given as 51% against published estimates of ~25%–28%.
7.	Could a patient with endometrial cancer receive fertility sparing treatment?	0.8	
8.	What options are available for fertility-sparing in patients with endometrial cancer?	0.4	Not mentioned: Ovarian transposition is an option for cervical cancer, not for endometrial cancer.
9.	Which progesterone is best for the treatment of endometrial cancer?	1	
10.	What are predictors of response to progesterone for treatment of endometrial cancer?	0.8	
11.	What is the exact duration of therapeutic benefit from progesterone therapy in patients with endometrial cancer wishing to preserve fertility?	0.2	The answer highlighted factors related to treatment response and treatment process, rather than giving evidence-based estimate of duration.
12.	Is pregnancy possible after receiving progesterone therapy for endometrial cancer?	0.8	
13.	What are the risks of fertility preservation in endometrial cancer?	0.4	Not mentioned: Lack of specimen may limit detection of Lynch syndrome and synchronous ovarian cancer.
14.	What is the prognosis of endometrial cancer?	0.4	Risk groups were not included in the response.
15.	What is the standard primary surgical management for endometrial cancer?	0.4	Not mentioned: Management is based on stage and grade. e.g., omentectomy is indicated for certain histological types, even if there is no detectable omental disease.
16.	What is the best surgical approach for management of endometrial cancer?	1	
17.	What is the role of surgery in patients with advance stage endometrial cancer?	0.6	
18.	I am recently diagnosed with metastatic endometrial cancer and my doctor is planning for surgery. Do I need any further treatment after surgery?	0	Not mentioned: Adjuvant chemotherapy is used for localized disease. Metastatic disease it treated with palliative chemotherapy. Radiotherapy is not used for metastatic disease.
19.	I am recently diagnosed with metastatic endometrial cancer and my doctor decided that I'm not fit for surgery. What are the treatment options?	-0.2	Not specified that radiotherapy, especially brachytherapy, is used for metastatic endometrial cancer.
20.	What are the treatment options for recurrence of endometrial cancer?	0.2	Not mentioned: Palliative chemotherapy with or without immune checkpoint inhibitors should be first option. PI3K/AKT/mTOR pathway inhibitors are not approved for endometrial cancer.

MMR: mismatch repair; TVUS: transvaginal ultrasound; PI3K: phosphatidylinositol 3-kinase; AKT: protein kinase B; mTOR: mammalian target of rapamycin.

irrelevant). These deficient responses were related to questions regarding treatment (Q 4, 11, 12, and 14–18 for cervical cancer, and Q 8, 11, 13–15, and 18–20 for endometrial cancer), screening (Q 6 and 7 for cervical cancer), and risk factors (Q1 for endometrial cancer).

Responses to questions about prevention, eligibility for screening, premalignant lesions of the cervix, and the side effects of treatment of cervical cancer were graded 'accurate.' For endometrial cancer, responses on genetic mutations, genetic testing, lifetime risk of developing uterine cancer, and non-surgical treatment were rated 'accurate.' However, only 6/20 (30.0%) responses for cervical cancer, and 2/20 (10.0%) responses for endometrial cancer received a maximal mean score of 1. More than 85% of responses were deemed appropriate and consistent by at least two reviewers.

Generally, mean score was higher for questions related to prevention and risk factors for cervical cancer, a pattern is consistent with prior studies evaluating AI-generated responses to questions related to hepatocellular carcinoma[11] and prostate cancer.[12]

However, methodological differences exist between studies. For example, in a similar study, ChatGPT's answers to common questions on colon cancer were judged in relation to the handouts of the American Society of Colon and Rectal Surgeons.[13] Their scoring system also differed from ours.

According to a study, AI chatbots scored above pass-mark for the United States Medical Licensing Examination[18] and made correct emergency room diagnoses more often than the human doctors did (97% vs. 87%).[19] However, in our study, AI responses related to stage-specific treatment recommendations were generally inadequate. These included irrelevant or incomplete information, such as the failure to mention hysterectomy for CIN3 for women who have completed childbirth, or omitting to mention surgical risks in cervical cancer and importance of combined positive score in selecting patients for immune checkpoint inhibitor in metastatic or recurrent cancer of cervix.

There are a few limitations to this study. First, it evaluated a single version of ChatGPT (version 3.5) at a specific time point, generating a static 'snapshot' of a rapidly evolving platform. Second, the evaluation was based on a fixed set of questions. Future studies could use varied question phrasing (reflecting how patients and nonspecialist healthcare personnel might ask questions) at different time points to better capture AI evolution across diverse cultural and linguistic contexts. This approach may also help AI systems better interpret users' intentions across cultures and languages. Third, our questions were generated in English, though our patients and many general practitioners in Oman may ask questions in Arabic as well as in English. Therefore, future regional studies could include question sets in Arabic. Finally, we used a three-point scoring system (+1, 0, or -1), which may have introduced some subjectivity in response categorization, although all reviewers were required to provide explanations for their ratings.

The public use of AI platforms for medical information will continue to expand, whether health authorities approve of it or not. Newer generations of large language models, already far more capable than the version evaluated in this study, are likely to become increasingly accurate, better aligned with medical knowledge, and more consistent over time. Independent safety organizations, including Emergency Care Research Institute, have identified misuse of AI chatbots as a major health technology hazard, largely because these systems are known to confidently dispense incorrect or misleading information.[20]

Another emerging risk is the proliferation of unregulated or poorly governed AI chatbots. Many users lack the skills to distinguish high-quality sources from unreliable or predatory ones. Addressing this gap will require education imparted early in life, alongside clearer standards for how AI systems should present health related information with high-quality citations. Future research should therefore focus on the safe integration of AI into healthcare, with awareness and skepticism from both clinicians and public, rather than treating AI as an 'always-dependable' source of medical advice.

CONCLUSION

This study assessed the accuracy of information generated by ChatGPT 3.5 on prevention, screening, diagnosis, and stage-specific treatment of cervical and uterine cancers. While ChatGPT provided accurate information concerning prevention and screening, responses related to treatment were often inadequate, and occasionally inaccurate. These findings represent an evident limitation for relying on AI-generated models in clinical decision-making and emphasize a clear need for cautious practice of these tools, with appropriate human oversight aligned with clinical guidelines.

Future studies should compare the performance of newer iterations of ChatGPT with other AI tools (Claude, Gemini, CoPilot, Grok, Perplexity, DeepSeek, Qwen, etc.) in generating evidence-based medical information for both medical personnel and patients.

Finally, researchers in Arabic speaking regions, including the Gulf Cooperation Council, should collaborate to develop comprehensive frameworks for integrating AI technologies and tools into clinical workflows that address our regional cultural and linguistic needs.

Disclosure
The authors declare no conflicts of interest. This study was supported by Sultan Qaboos University's grant IG/EPS/INFS/25/02.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 May;71(3):209-249.
2. Godfrey K, Agatha T, Nankumbi J. Breast cancer knowledge and breast self-examination practices among female university students in Kampala, Uganda: a descriptive study. Oman Med J 2016 Mar;31(2):129-134.
3. Kakotkin VV, Semina EV, Zadorkina TG, Agapov MA. Prevention strategies and early diagnosis of cervical cancer: current state and prospects. Diagnostics (Basel) 2023 Feb;13(4):610.
4. Shaffer KM, Turner KL, Siwik C, Gonzalez BD, Upasani R, Glazer JV, et al. Digital health and telehealth in cancer care: a scoping review of reviews. Lancet Digit Health 2023 May;5(5):e316-e327.
5. Senbekov M, Saliev T, Bukeyeva Z, Almabayeva A, Zhanaliyeva M, Aitenova N, et al. The recent progress and applications of digital technologies in healthcare: a review. Int J Telemed Appl 2020 Dec;2020(1):8830200.
6. Kashoub M, Al Abdali M, Al Shibli E, Al Hamrashdi H, Al Busaidi S, Al Rawahi M, et al. Artificial intelligence in medicine: a double-edged sword or a pandora's box? Oman Med J 2023 Sep;38(5):e542.
7. Perera Molligoda Arachchige AS. Role of ChatGPT 3.5 in emergency radiology, with a focus on cardiothoracic emergencies: proof with examples. iRadiology 2024;1:3.
8. Uprety D, Zhu D, West HJ. ChatGPT-A promising generative AI tool and its implications for cancer care. Cancer 2023 Aug;129(15):2284-2289.
9. Chen S, Kann BH, Foote MB, Aerts HJ, Savova GK, Mak RH, et al. The utility of ChatGPT for cancer treatment information. MedrXiv 2023 Mar 23:2023-03.
10. Johnson SB, King AJ, Warner EL, Aneja S, Kann BH, Bylund CL. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. JNCI Cancer Spectr 2023 Mar;7(2):pkad015.
11. Cao JJ, Kwon DH, Ghaziani TT, Kwo P, Tse G, Kesselman A, et al. Accuracy of information provided by ChatGPT regarding liver cancer surveillance and diagnosis. AJR Am J Roentgenol 2023 Oct;221(4):556-559.
12. Coskun B, Ocakoglu G, Yetemen M, Kaygisiz O. Can ChatGPT, an artificial intelligence language model, provide accurate and high-quality patient information on prostate cancer? Urology 2023;180:35-58.
13. Emile SH, Horesh N, Freund M, Pellino G, Oliveira L, Wignakumar A, et al. How appropriate are answers of online chat-based artificial intelligence (ChatGPT) to common questions on colon cancer? Surgery 2023 Nov;174(5):1273-1275.
14. Wei K, Fritz C, Rajasekaran K. Answering head and neck cancer questions: an assessment of ChatGPT responses. Am J Otolaryngol 2024;45(1):104085.
15. Alasker A, Alsalamah S, Alshathri N, Almansour N, Alsalamah F, Alghafees M, et al. Performance of large language models (LLMs) in providing prostate cancer information. BMC Urol 2024 Aug;24(1):177.
16. Freire Y, Santamaría Laorden A, Orejas Pérez J, Gómez Sánchez M, Díaz-Flores García V, Suárez A. ChatGPT performance in prosthodontics: assessment of accuracy and repeatability in answer generation. J Prosthet Dent 2024 Apr;131(4):659.e1-659.e6.
17. Olszewski R, Brzezinski J, Watros K, Manczak M, Owoc J, Jeziorski K. Exploring the role of AI-driven chatbots in patient care: a critical evaluation amidst healthcare staff shortages. Eur Heart J 2024 Oct;45(Supplement_1):eh ae666-3495.
18. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature 2023 Aug;620(7972):172-180.
19. Berg HT, van Bakel B, van de Wouw L, Jie KE, Schipper A, Jansen H, et al. ChatGPT and generating a differential diagnosis early in an emergency department presentation. Ann Emerg Med 2024 Jan;83(1):83-86.
20. ECRI. Misuse of AI chatbots tops annual list of health technology hazards. ECRI News. 2026 [cited 2026 Jan 24]. Available from: https://home.ecri.org/blogs/ecri-news/misuse-of-ai-chatbots-tops-annual-list-of-health-technology-hazards.