

Analysing Usability of ChatGPT-4 as a Source of Information about Human Papilloma Virus Vaccines

Çağdaş Demiroğlu

Medical Point Gaziantep Hospital, Clinic Obstetrics and Gynecology, Gaziantep, Turkey

ABSTRACT

Purpose: Human papillomavirus (HPV) is a highly prevalent sexually transmitted pathogen, with nearly 80% of sexually active individuals becoming infected during their lifetime. Although recently developed vaccines provide effective protection, increasing reliance on internet-based and artificial intelligence (AI)-generated medical information raises concerns regarding the accuracy and adequacy of such content. This study aimed to evaluate the accuracy, quality, and readability of HPV vaccine-related information generated by ChatGPT.

Methods: The 25 most-searched for HPV vaccine-related keywords were identified through Google Trends. ChatGPT responses to these queries were collected and assessed using the Ensuring Quality Information for Patients (EQIP) criteria, the Flesch-Kincaid Grade Level (FKGL), the Flesch-Kincaid Reading Ease (FKRE), and two Likert scales (3-point and 5-point). Evaluations were independently performed by two healthcare professionals. Statistical comparisons were made across EQIP categories and readability indices.

Results: The three most frequently searched phrases were “vaccine for HPV,” “HPV vaccine side effects,” and “HPV side effects.” The mean EQIP score was 62.48, the mean FKRE score was 44.89, and the mean FKGL level was 12.08. The mean scores on the 5-point and 3-point Likert scales were 4.12 and 2.2, respectively. No statistically significant differences were observed between EQIP categories ($p=0.332$) or in FKGL and FKRE comparisons ($p=0.244$ and $p=0.157$).

Conclusion: ChatGPT provided generally satisfactory information regarding HPV vaccines; however, several quality limitations were identified. The content demonstrated adequate scientific accuracy but required a reading level consistent with high school education. EQIP scoring indicated that the information was of “good quality with minor issues”. In particular, simplifying technical language, improving structural organization, and incorporating more patient-centered explanations may substantially increase the accessibility and practical value of AI-generated medical content.

Keywords: Artificial intelligence, papillomavirus vaccines, patient education as topic, information dissemination

INTRODUCTION

Human papillomavirus (HPV) is a common sexually transmitted pathogen and the leading etiological factor for cervical cancer. According to the World Health Organization, approximately 80% of sexually active individuals will contract HPV at some point in their lives. Globally, an estimated 570,000 women and 60,000 men develop HPV-related cancers each year.¹

HPV infection is often asymptomatic, though mild symptoms may occur in symptomatic cases.² Vaccines developed in recent years provide effective protection against HPV infection,

playing a crucial role in preventing genital warts, cervical cancer, and other precancerous lesions.³ In countries with vaccination programs, a significant reduction in precancerous lesions and genital warts has been observed, correlated with high vaccination rates.⁴ Despite these proven benefits, concerns and misconceptions surrounding HPV vaccines persist.⁵ Misinformation, often fueled by social media, contributes to vaccine hesitancy in certain communities.⁶

ChatGPT, an artificial intelligence (AI) chatbot developed by OpenAI, has gained popularity for its ability to generate detailed responses to user queries. Using deep learning techniques,



Address for Correspondence: Çağdaş Demiroğlu, Medical Point Gaziantep Hospital, Clinic Obstetrics and Gynecology, Gaziantep, Turkey

E-mail: cagdasdemiroglu@gmail.com **ORCID ID:** orcid.org/0000-0002-7011-3890

Received: 24.11.2025 **Accepted:** 06.12.2025 **Epub:** 30.01.2026

Cite this article as: Demiroğlu Ç. Analysing usability of ChatGPT-4 as a source of information about human papilloma virus vaccines. Anat J Obstet Gynecol Res. [Epub Ahead of Print]



Copyright© 2025 The Author(s). Published by Galenos Publishing House on behalf of National Society of Gynecology and Obstetrics. This is an open access article under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

specifically a variant of the transformer-based neural network, it is capable of producing contextually relevant and human-like responses. Its accessibility and rapid response generation have attracted millions of users worldwide.⁷ In terms of public health information about HPV, the availability of clear, accurate information is clearly important and ChatGPT is currently a major source of widely used information about these vaccines.

Although clinicians largely agree on the safety and efficacy of HPV vaccines, access to accurate information remains a challenge for the general population. Misinformation can lead to patients downplaying risks or neglecting preventive measures.⁸ With the increasing use of the internet and AI-based platforms, the accessibility of information has improved, but concerns about its accuracy and adequacy remain. While patients who consult clinicians can receive accurate information, the accuracy and readability of information from internet searches and AI platforms, such as ChatGPT, have not been extensively reported in the literature.⁹

The aim of this study was to evaluate the accuracy and readability of information about HPV vaccines provided by ChatGPT. Understanding this will help assess the potential of widely accessible information to mislead patients and inform strategies to address misinformation.

METHODS

This study was conducted on October 1, 2024, at Gaziantep Medical Point Hospital. As the study did not involve human subjects, patient data, or any *in vivo* procedures, approval from an Institutional Review Board was not required. The research was performed in accordance with the principles of the Declaration of Helsinki. The most frequently searched HPV vaccine-related keywords were identified using Google Trends (<https://trends.google.com/>). The search was performed on October 1, 2024, with the region set to “worldwide” and all available data from January 2004 to the date of access included. Prior to conducting the search, all browser data were cleared to minimize potential personalization bias. From the generated list, the top 25 most relevant and frequently searched queries in English were selected, excluding repetitive, irrelevant, or non-English terms. All available data from 2004 to the present were included, and the search region was set to “worldwide”.

The 25 selected keywords identified via Google Trends were used as direct user queries and entered verbatim into the ChatGPT interface. Each keyword was submitted as a standalone query, without additional contextual framing, prompts, or follow-up questions, in order to reflect typical real-world patient search behavior. Each query was entered in a new and independent chat session to prevent contextual carryover and response contamination. All questions and responses were categorized into six groups: Condition or illness, medication or product, prevention or aftercare, test, operation, and investigation or procedure. A formal sample size calculation was not applicable because the dataset consisted of the top 25 globally most-searched HPV vaccine-related keywords identified via Google Trends. The number of items was therefore determined by the natural structure of the dataset rather than by investigator selection, which is

methodologically standard in studies analyzing search-trend-based query lists.

To evaluate the quality of the responses, a Google Form containing 20 items from the Ensuring Quality Information for Patients (EQIP) checklist was used. This study used the original 20-item EQIP version developed by Moult et al.¹⁰, which is the shorter, preliminary validation form, rather than the extended 36-item version used in later adaptations. Each item was scored as “yes” (1 point), “partially” (0.5 points), “no” (0 points), or “not applicable”. The EQIP score was calculated as the sum of the “yes” responses divided by the total number of applicable items (adjusted for “not applicable” responses). Scores were presented as percentages and categorized into four groups: 76-100% (well-written, high-quality); 51-75% (good quality with minor problems); 26-50% (serious problems with quality); and 0-25% (significant quality issues).¹⁰

Readability was assessed using two parameters: the Flesch-Kincaid Reading Ease score (FKRE) and the Flesch-Kincaid Grade Level score (FKGL). FKRE was calculated using the standard formula $206.835 - [1.015 \times \text{average sentence length (ASL)}] - [84.6 \times \text{average syllables per word (ASW)}]$. The score indicated the educational level required for comprehension, with scores below 30 indicating university-level comprehension. FKGL was calculated using the standard formula $(0.39 \times \text{ASL}) + (11.8 \times \text{ASW}) - 15.59$, where higher values correspond to more complex text and a higher required grade level. A lower FKGL score indicates easier comprehension.¹¹ In addition, information accuracy and completeness were evaluated using both 3-point and 5-point Likert scales. For accuracy, a 5-point scale was used (1= “very low accuracy with serious errors,” 5= “very high accuracy with no errors”). For completeness, a 3-point scale was used (1= “incomplete,” 3= “comprehensive coverage”). For reproducibility, “accuracy” was operationally defined as the degree to which ChatGPT responses aligned with current evidence-based medical guidelines and contained no factual errors or misleading statements. “Completeness” was defined as the extent to which responses addressed all major components of the queried topic, including definition, causes, symptoms, prevention, management, and when applicable, treatment alternatives.¹² All evaluations were conducted by two independent healthcare professionals (C.D. and I.T.S.) to minimize bias, both of whom are board-certified physicians, with C.D. specializing in obstetrics and gynecology and I.T.S. specializing in general surgery.

Statistical Analysis

Quantitative variables were expressed as mean \pm standard deviation (SD). The Shapiro-Wilk test was used to assess the normality of continuous variables prior to selecting appropriate statistical tests. The Kruskal-Wallis H test was used to compare EQIP, FKRE, FKGL, and Likert scores across more than two independent groups as this rank-based nonparametric test compares the distribution of median ranks rather than means. Bonferroni post-hoc correction was applied for multiple group comparisons. Correlations between numerical variables were assessed using the nonparametric Spearman’s rank correlation test, as the data did not follow a normal distribution. Statistical significance was set at $p < 0.05$ for all tests. All analyses were

performed using IBM SPSS Statistics for Windows, version 21.0 (IBM Corp., Armonk, NY, USA).

RESULTS

The three most frequently searched phrases were “vaccine for HPV,” “HPV vaccine side effects,” and “HPV side effects”. The full list of the 25 most used phrases is provided in Table 1. The top three countries where HPV vaccine searches were most frequent were Hong Kong, Taiwan, and Singapore (Figure 1), with the top 10 countries listed in Table 2.

Table 3 presents the minimum, maximum, means, and SDs of the EQIP, FKRE, and FKGL scores. The EQIP scores for the texts ranged from 50 to 72, with a mean score of 62.48. The FKRE scores varied between 17.6 and 64.2, with a mean score of 44.89. A lower FKRE score indicates increased reading difficulty; accordingly, the mean FKRE score of 44.89 corresponds to a reading level generally considered difficult and consistent with high school-level comprehension rather than university-level understanding. The FKGL scores ranged from 8.4 to 21.1, with a mean of

12.08. Higher FKGL values similarly correspond to more complex text, reflecting a reading level typically consistent with high school education or above.

On the 5-point Likert scale, the texts received scores ranging from 3 to 5, with an average of 4.12. On the 3-point Likert scale, scores ranged from 1 to 3, with an average score of 2.2.

The texts generated by ChatGPT were categorized into six groups (Figure 2) for comparison. No statistically significant differences were observed between these categories in terms of EQUIP scores ($p=0.332$). Similarly, no significant differences were found between the categories for FKGL ($p=0.244$) or FKRE ($p=0.157$) scores.

DISCUSSION

The findings of this study indicate that while the content of the texts generated by ChatGPT was generally satisfactory, there were notable deficiencies in terms of quality. The scientific content was positively rated according to the Likert scales, with average scores of 2.2 on the 3-point scale and 4.12 on the 5-point scale. Furthermore, the FKGL and FKRE readability

Table 1. The top 25 keywords queried globally for HPV vaccines between 2004 and 2024, according to Google Trends data and EQIP classifications

	Questions	Popularity percentage	Classification of the topic according to EQIP Prevention or after care
1	Vaccine for HPV	100	Prevention or aftercare
2	HPV vaccine side effect	47	Prevention or aftercare
3	HPV side effects	46	Condition or illness
4	What is HPV	45	Condition or illness
5	What is HPV vaccine	45	Condition or illness
6	HPV cancer	43	Condition or illness
7	Cancer vaccine	42	Condition or illness
8	Gardasil	33	Medication or product
9	Gardasil HPV vaccine	33	Medication or product
10	Gardasil vaccine	31	Medication or product
11	HPV vaccine men	31	Prevention or aftercare
12	HPV vaccine age	31	Prevention or aftercare
13	HPV men	29	Prevention or aftercare
14	HPV vaccine schedule	26	Test, operation, investigation or procedure
15	Vaccine schedule	25	Test, operation, investigation or procedure
16	Warts	25	Condition or illness
17	HPV warts	24	Condition or illness
18	Cervical cancer vaccine	23	Medication or product
19	HPV vaccine cervical cancer	23	Medication or product
20	Cervical cancer	22	Condition or illness
21	What is the HPV vaccine	19	Medication or product
22	HPV vaccine symptoms	18	Condition or illness
23	HPV vaccine cost	18	Miscellaneous
24	Vaccines	17	Miscellaneous
25	HPV symptoms	17	Condition or illness

assessments yielded values indicating that the texts were relatively difficult to read, requiring at least a high school education level for adequate comprehension.

The three most frequently searched phrases in this study were “vaccine for HPV,” “HPV vaccine side effects,” and “HPV side effects,” suggesting that the public were particularly interested in the existence and safety of the HPV vaccine. Notably, the highest search activity was concentrated in small, largely ethnically Chinese Asian countries, with Hong Kong, Taiwan, and Singapore leading in searches. This geographic distribution may reflect regional differences in sexual health awareness,



Figure 1. Global search interest in HPV vaccines from 2004 to 2024, based on Google Trends data (regions with medium search volume not included)

HPV: Human papillomavirus

Table 2. Interest by region

	Country	Popularity percentage
1	Hong Kong	100
2	Taiwan	84
3	Singapore	68
4	Denmark	63
5	Austria	52
6	Norway	51
7	New Zealand	44
8	Ireland	42
9	United States	35
10	Japan	35

vaccination policies, and patterns of online information-seeking behavior.

ChatGPT has increasingly become a go-to resource for patients seeking information before consulting healthcare providers. Since 2004, the number of questions posed to chatbots has steadily increased (Figure 3). In the present study, the quality of the text responses was assessed using FKGL and FKRE scores. A study by Şahin et al.¹³ on erectile dysfunction found that the readability of chatbot-generated texts was low, generally requiring a sixth-grade reading level. In contrast, our results suggested that the FKGL scores corresponded to a high school reading level, making the content more challenging to understand. This is a significant finding, as AI platforms are used by a wide range of individuals, and greater readability would facilitate better comprehension. Poor readability has the potential to cause misunderstandings, which may result in treatment refusal or hesitancy.

In addition to readability, it is important that the information provided by ChatGPT is scientifically accurate. Accurate information can not only reduce the workload for healthcare professionals but also build patient trust in treatment protocols. A study by Goodman et al.¹⁴ in 2023 evaluated chatbot-generated answers on medical topics using two separate scales. More than 50% of the responses were rated as completely or nearly correct, with most considered comprehensive.¹⁴ In our study, two different healthcare professionals evaluated the texts using 3-point and 5-point Likert scales, and the scientific content was found to be of high quality. These findings align with previous research, suggesting that ChatGPT-generated responses are satisfactory in terms of content accuracy.

The EQIP scoring system is another valuable tool for evaluating the quality of health information. It assesses the clarity, structure, use of sources, accuracy, relevance, and comprehensiveness of patient education materials. A study by Walker et al.¹⁵ in 2023 used EQIP to evaluate ChatGPT-generated answers for benign and malignant liver tumors. The texts were rated as “low” to “medium quality”.¹⁵ Similarly, a study by Erden et al.¹⁶ in 2023 evaluated ChatGPT responses to osteoporosis-related queries and found the EQIP scores to reflect “medium quality”. In the present study, the average EQIP score was in the “good quality with minor problems” category. The slight improvement in EQIP scores compared to earlier studies may be attributed to improvements in chatbot performance and learning processes, as ChatGPT

Table 3. Descriptive statistics of EQIP, FKRE, FKGL, and Likert scores (mean ± SD, min-max)

	Minimum	Maximum	Mean	Std. deviation
Ensuring Quality Information for Patients score	50	72	62.48	±5.16
The Flesch-Kincaid Reading Ease score	17.6	64.2	44.89	±11.98
The Flesch-Kincaid Grade Level score	8.4	21.1	12.08	±2.85
3 points Likert scale	1	3	2.2	±0.65
5 points Likert scale	3	5	4.12	±0.73

EQIP: Ensuring Quality Information for Patients, FKRE: Flesch-Kincaid Reading Ease, FKGL: Flesch-Kincaid Grade Level, SD: Standard deviation, min-max: Minimum-maximum

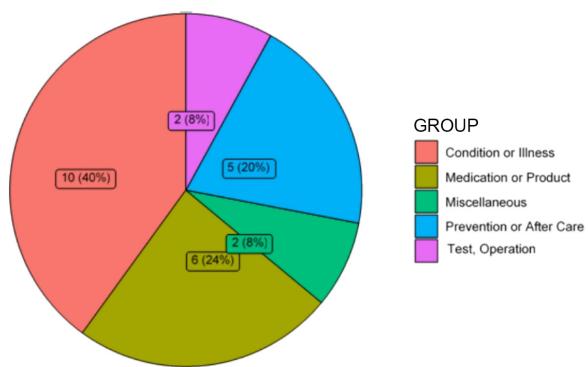


Figure 2. The percentage of text categories based on the EQIP
EQIP: Ensuring Quality Information for Patients

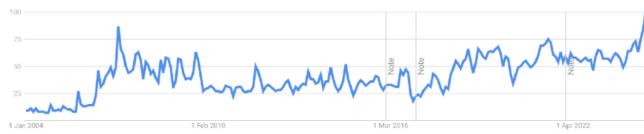


Figure 3. Global search interest based on Google Trends data from 2004 to 2024

has undergone several revisions and the number of queries posed to AI has increased over time. However, despite this improvement, the scientific quality of the content remains below the desired level.

Study Limitations

A limitation of this study is the reliance on a single chatbot platform. Results may differ when using other AI models. Furthermore, the quality and readability of the responses may vary depending on the specific version of the AI used. As AI continues to evolve, the accuracy, reliability, and readability of such texts are expected to improve. Another limitation is that the study was conducted solely in English; future studies in other languages may yield different results.

Recent studies published in 2023-2024 similarly emphasize the need for improving the clarity and patient-oriented design of AI-generated health information, highlighting that readability and quality scores often remain suboptimal despite high factual accuracy.^{12,13,16}

CONCLUSION

The accessibility of accurate and high-quality information on important public health topics, such as HPV vaccines, is essential. AI-based chatbots, like ChatGPT, can play a valuable role in providing such information. These platforms have the potential to improve patient education and foster confidence in treatment. This study demonstrated that ChatGPT is an informative and easily accessible source for HPV vaccine information in the English language. However, the findings also highlight that the content provided still falls short in terms of comprehensiveness and readability. In particular, simplifying technical language, improving structural organization, and

incorporating more patient-centered explanations may substantially increase the accessibility and practical value of AI-generated medical content.

Ethics

Ethics Committee Approval: As the study did not involve human subjects, patient data, or any *in vivo* procedures, approval from an Institutional Review Board was not required.

Informed Consent: Not applicable, as this study did not involve human participants, patient data, or identifiable personal information.

Acknowledgement

The author would like to thank İbrahim Tayfun Şahiner for his assistance with the statistical analysis.

Footnotes

Financial Disclosure: The author declared that this study received no financial support.

REFERENCES

1. Kothari A. Human papilloma virus (HPV) and cervical cancer. Independent Nurse. 2020;2020(2):20-22.
2. Aydin TÖ, Ceylan Y, Aydin DE, Sarac ÖD, Güraslan H. Could clinically suspicious cervix predict cervical premalignant and malignant lesions in postmenopausal women?. Anat J Obstet Gynecol Res. 2024;1(2):61-65.
3. Garland SM, Hernandez-Avila M, Wheeler CM, et al. Quadrivalent vaccine against human papillomavirus to prevent anogenital diseases. N Engl J Med. 2007;356(19):1928-1943.
4. Drolet M, Bénard É, Pérez N, Brisson M; HPV vaccination Impact Study Group. Population-level impact and herd effects following the introduction of human papillomavirus vaccination programmes: updated systematic review and meta-analysis. Lancet. 2019;394(10197):497-509.
5. Szilagyi PG, Albertin CS, Gurfinkel D, et al. Prevalence and characteristics of HPV vaccine hesitancy among parents of adolescents across the US. Vaccine. 2020;38(38):6027-6037.
6. Jennings W, Stoker G, Bunting H, et al. Lack of trust, conspiracy beliefs, and social media use predict COVID-19 vaccine hesitancy. Vaccines (Basel). 2021;9(6):593.
7. Xue VW, Lei P, Cho WC. The potential impact of ChatGPT in clinical and translational medicine. Clin Transl Med. 2023;13(3):e1216.
8. Dubé E, Gagnon D, MacDonald NE; SAGE Working Group on Vaccine Hesitancy. Strategies intended to address vaccine hesitancy: review of published reviews. Vaccine. 2015;33(34):4191-4203.
9. van der Meer TGLA, Jin Y. Seeking formula for misinformation treatment in public health crises: the effects of corrective information type and source. Health Commun. 2020;35(5):560-575.
10. Moult B, Franck LS, Brady H. Ensuring quality information for patients: development and preliminary validation of a new instrument to improve the quality of written health care information. Health Expect. 2004;7(2):165-175.
11. Boles CD, Liu Y, November-Rider D. Readability levels of dental patient education brochures. J Dent Hyg. 2016;90(1):28-34.
12. Musheyev D, Pan A, Kabarriti AE, Loeb S, Borin JF. Quality of information about kidney stones from artificial intelligence chatbots. J Endourol. 2024;38(10):1056-1061.
13. Şahin MF, Ateş H, Keleş A, et al. Responses of five different artificial intelligence chatbots to the top searched queries about erectile

dysfunction: a comparative analysis. *J Med Syst.* 2024;48(1):38.

14. Goodman RS, Patrinely JR, Stone CA Jr, et al. Accuracy and reliability of chatbot responses to physician questions. *JAMA Netw Open.* 2023;6(10):e2336483.

15. Walker HL, Ghani S, Kuemmerli C, et al. Reliability of medical information provided by ChatGPT: assessment against clinical guidelines and patient information quality instrument. *J Med Internet Res.* 2023;25:e47479.

16. Erden Y, Temel MH, Bağcier F. Artificial intelligence insights into osteoporosis: assessing ChatGPT's information quality and readability. *Arch Osteoporos.* 2024;19(1):17.