Quality Assessment of AI-Generated Response to Patient-Reported Information Gaps Regarding HPV-Mediated Oropharyngeal Cancer

Patrick Bidros[1], Clayton Bobo[1], Anthony Mahairas[2] and Melina Windon[2,3]
[1]University of Kentucky College of Medicine, Lexington, KY, 40509
[2]University of Kentucky College of Medicine, Department of Otolaryngology, Lexington, KY 40509
[3]Markey Cancer Center, University of Kentucky, Lexington, KY

BACKGROUND

- Human papillomavirus (HPV)-mediated oropharyngeal cancer (HPV-OPC) has been rising in incidence in the United States and worldwide.[1]
- Even among patients with HPV-OPC, many remain under-informed: among a cohort of 300 patients, less than half were able to correctly answer questions about HPV. [2,3]
- Previously described specific knowledge gaps among patients include the HPV tumor status of their cancer, misinformation or uncertainty about the past and future transmission of HPV, and outcomes of treatment of their OPC.[4]
- Patients frequently turn to the internet to fill these gaps, especially for sensitive information.[5,6]
- AI tools (e.g., ChatGPT, Google AI) are now commonly integrated into search engines, yet the accuracy and understandability of responses remain unclear.
- This study evaluates AI-generated information for patient education about HPV-OPC.

METHODS AND MATERIALS

Patients with HPV-OPC treated at the NCI-Designated Markey Cancer Center in Lexington, KY from 06/05/2021 – 12/31/2023 were contacted and verbally consented by telephone to participate in a phone interview.
Responses were summarized into colloquial text and used to prompt two AIs (ChatGPT (OpenAI, GPT-4 model, March 2023 version) and Gemini (Google DeepMind, Gemini 1.5 model) on 10/18/2024. Reading level was assessed using Flesch Kincaid Reading Ease. The responses from the AI models were then distributed to board-certified head and neck oncologists using REDcap and reviewed in 6 domains (accuracy, clarity, relevance, completeness, provision of resources and references, and usefulness) using the validated Quality Assessment of Medical Artificial Intelligence (QAMAI) tool.[7] Oncologists were asked additional open-ended questions about the output quality. Results were compared using the Mann Whitney-U test. This study was reviewed and approved by the University of Kentucky Institutional Review Board.

RESULTS

- The following prompts were generated in layperson terms: “pros and cons getting treatment for hpv throat cancer”, “side effects radiation hpv throat cancer”, “treatment options hpv throat cancer”, “would the vaccine have stopped me from getting hpv throat cancer”, “what is hpv”, and “how is hpv spread.” These prompts were then verified with the remaining interviews, and no additional themes emerged.
- Out of 192 participants contacted, 43 consented and completed the interview.
- Reading level was college level (55.0) for Google AI responses, whereas for ChatGPT4, it was 10th-12th grade level (44.9).
- From the QAMAI analysis, ChatGPT4 scored an average of 20.5/30, demonstrating fair quality, and GoogleAI scored an average of 22.5/30, demonstrating good quality.
- Moreover, the QAMAI score was significantly better for Google AI ($p<0.001$), primarily due to the provision of resources.
- Most (5/6) physicians found that the information was over-simplified or misleading, pointing to the lack of priority to side effects and treatments.
- Other physicians pointed out that some of the treatment options mentioned were either outdated or presented in a fashion that led the reader to assume the adjuvant modality discussed was a stand-alone treatment instead of adjuvant therapy.

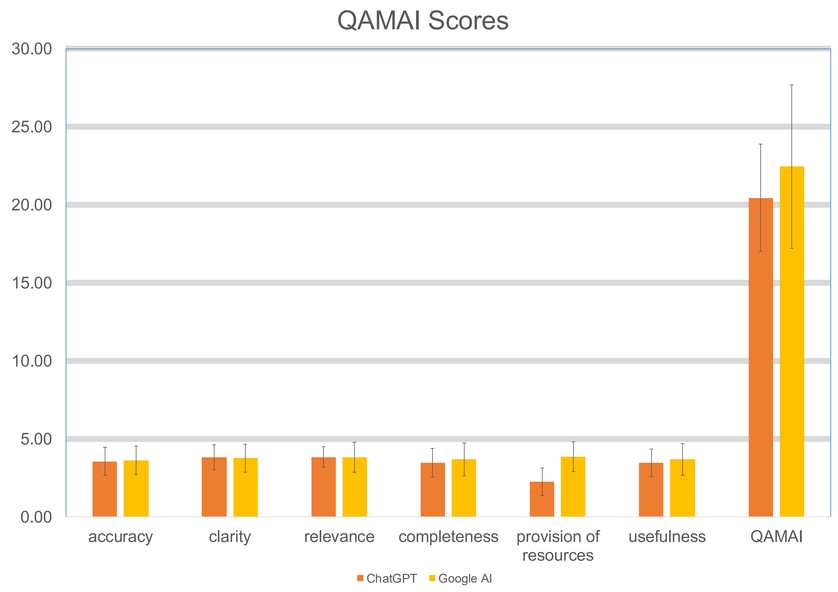

Figure 1: QAMAI Scores of ChatGPT and GoogleAI in 6 respective domains and overall score

Themes:	Physician Comments:
Misinformation	• “Equally weighted treatment benefits (i.e. cure) with side effects, suggesting a 50-50 situation of death from one or life with the other” • “This answer has some inaccuracies. EBRT is fairly outdated. Neck dissection is almost always indicated without clinical disease (elective ND)”
Complexity	• “Elevated language that will be difficult for patient to understand” • “It is an oversimplification and may be misleading for the patient as this general information may not be applicable to the patient's specific disease”
Lack of References	• “No references and no data were provided to substantiate statements”
Educational Merit	• “Good things to consider, covers the basics” • “Good comprehensive umbrella overview”

Table 1: Themes of misinformation, complexity, lack of references but also educational merit emerged among physicians reviewing AI-generated answers (above)

CONCLUSION

- AI-generated information regarding HPV-positive oropharyngeal cancer is overall good.
- Despite this, it can provide misinformation and reading level is too complex for most patients
- We found that the particular knowledge gaps that HPV-OPC may seek to fill include treatment side effects, general knowledge about HPV, and HPV transmission.
- As AI models evolve, it is essential that physicians remain the ‘human in the loop’ to build high quality educational sources and advise patients

ACKNOWLEDGMENTS

PSMRF: The project described was supported by the National Center for Advancing Translational Sciences, through Grant UL1TR001998. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES

1. Wu J, Xiao F, Zheng Y, Lin Y, Wang HL. Worldwide trend in human papillomavirus-attributable cancer incidence rates between 1990 and 2012 and Bayesian projection to 2030. *Cancer*. 2021;127(17):3172-3182. doi:10.1002/cncr.33628
2. Inglehart RC, Taberna M, Pickard RKL, et al. HPV knowledge gaps and information seeking by oral cancer patients. *Oral Oncol*. 2016;63:23-29. doi:10.1016/j.oraloncology.2016.10.021
3. Torabi SJ, Kasle DA, Su-Velez BM, et al. A 2020 Update on Public Awareness of Head and Neck Cancers. *Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg*. 2022;166(2):305-312. doi:10.1177/01945998211006932
4. Milbury K, Rosenthal DI, El-Naggar A, Badr H. An exploratory study of the informational and psychosocial needs of patients with human papillomavirus-associated oropharyngeal cancer. *Oral Oncol*. 2013;49(11):1067-1071. doi:10.1016/j.oraloncology.2013.07.010
5. Baxi SS, Shuman AG, Corner GW, et al. Sharing a diagnosis of HPV-related head and neck cancer: the emotions, the confusion, and what patients want to know. *Head Neck*. 2013;35(11):1534-1541. doi:10.1002/hed.23182
6. Windon MJ, Le D, D'Souza G, et al. Treatment decision-making among patients with oropharyngeal squamous cell cancer: A qualitative study. *Oral Oncol*. 2021;112:105044. doi:10.1016/j.oraloncology.2020.105044
7. Vaira LA, Lechien JR, Abbate V, et al. Validation of the Quality Analysis of Medical Artificial Intelligence (QAMAI) tool: a new tool to assess the quality of health information provided by AI platforms. *Eur Arch Oto-Rhino-Laryngol Off J Eur Fed Oto-Rhino-Laryngol Soc EUFOS Affil Ger Soc Oto-Rhino-Laryngol - Head Neck Surg*. 2024;281(11):6123-6131. doi:10.1007/s00405-024-08710-0