Poster 12: Comparative evaluation of artificial intelligence-generated responses to patient questions on HPV and cervical cancer

Presenting Author: Emma Barr, MD, Adventist Health White Memorial

Topic
Cervical

Objectives
To evaluate the quality of publicly available artificial intelligence (AI) assistant-generated responses to common patient questions regarding HPV and cervical cancer.

Methods
Three free generative AI assistants (ChatGPT, Copilot, Gemini) were queried with three questions: “If I have HPV, will I get cervical cancer?”, “Is the HPV vaccine safe?”, and “Does the HPV vaccine prevent cervical cancer?” Each assistant was queried three times for each of the three questions, each time in a separate private browsing session to account for response variability. One response per question per assistant was randomly selected (nine total), de-identified, and presented in random order in an anonymous survey. The survey asked participants to evaluate each AI response across six domains: relevance, accuracy, completeness, clarity, coherence, and appropriateness using a 4-point Likert scale. Each response was evaluated for its potential to cause harm if relied upon without physician input (yes/no). The survey was distributed to gynecologic oncology (GO) attendings and fellows, and obstetrics and gynecology (OBGYN) attendings and residents. For each AI assistant, a composite quality score was calculated as the median across all domains and all three questions.

Results
Eight clinicians responded to the survey, including one GO attending, one GO fellow, four OBGYN attendings, and two OBGYN residents. Composite scores for ChatGPT, Copilot, and Gemini were 4.00, 3.75, and 4.00, respectively. The number of responses flagged as potentially harmful if relied upon without physician input from ChatGPT, Copilot, and Gemini were 5 (20.8%), 5 (20.8%), and 4 (16.7%), respectively.

Conclusions
Free, publicly available AI assistants generated responses to common HPV and cervical cancer questions that were rated generally high quality across multiple domains by clinicians. However, a meaningful proportion of responses were still perceived as potentially harmful if used without physician input. Given the increasing global reliance on digital health information, these findings highlight both the potential of AI to expand access to patient education and the need for careful oversight to ensure accuracy, safety, and equity in information delivery.