APPROPRIATENESS, COMPREHENSIVENESS, AND PHYSICIAN PERSPECTIVES ON CHATGPT'S RESPONSES TO COMMON QUESTIONS ABOUT IN VITRO FERTILIZATION

Authors: Thammatat Vorawandthanachai MD (1), Alexander Kotlyar MD (2)

Affiliations:
1. Department of Obstetrics and Gynecology, Maimonides Medical Center, Brooklyn, NY, USA
2. Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynecology, Maimonides Medical Center, Brooklyn, NY, USA

Background
ChatGPT, an artificial intelligence tool that integrates large data to generate human-like responses, is a popular tool with more than 1 billion monthly queries. Patients are increasingly utilizing ChatGPT to address health concerns, making it essential to evaluate the reliability of the information generated, as this tool can influence health-related decisions and overall well-being. However, investigation into the accuracy and comprehensiveness of the responses of ChatGPT has been limited, with available studies yielding inconclusive findings. Few studies have specifically examined the accuracy and comprehensibility of ChatGPT's responses regarding in vitro fertilization (IVF).

Objective
To investigate whether physicians in reproductive endocrinology and infertility (REI) assess ChatGPT's responses to FAQ's about IVF as accurate and comprehensible by patients and to determine physicians' favorability towards ChatGPT.

Materials and Methods
FAQ's about IVF were compiled from online discussion boards, and 9 most commonly asked questions were selected. These questions encompassed many topics including the definition of IVF, associated risks, duration, and inquiries related to embryo and genetic testing. The questions were queried into ChatGPT, and each question was repeated 3 times to ensure the validity of responses. The questions and ChatGPT's responses were then added to a questionnaire, and REI physicians from the US were asked to rank their accuracy and comprehensibility on a 5-point Likert scale. Physicians were blinded to the fact that these responses were generated by ChatGPT. Results were dichotomized as accurate or comprehensible (4-5) or not (1-3). Physicians then rated their attitudes toward ChatGPT (diagnostic accuracy, usefulness, educational utility, and IVF topic recommendation) on a 5-point Likert scale. Attitudes were dichotomized as positive (4-5) and negative (1-3). Descriptive analyses were recorded.

Results
Sixteen participants, who are REI physicians from private and public institutions in the US, completed the survey. Overall, ChatGPT's responses were rated as accurate (75-100%) and comprehensible (93.8-100%). Statements defining IVF and explanations of genetic testing in IVF were least ranked as accurate (75%), followed by risks and duration of IVF (87.5%), number of embryos used (93.8%), and age limit and rates of twin pregnancy in IVF (100%). ChatGPT's responses about age limit, rate of twins, and number of embryos used had the highest comprehensibility (100%) followed by definition of IVF, risks, duration, and genetic testing (93.8%). When assessing physicians' attitudes about ChatGPT, only 25% participants found ChatGPT to be an accurate tool, and 18.8% considered it better than other online resources.

However, 56.3% would recommend ChatGPT to supplement education, and 31.25% would recommend it for patient research on IVF topics.

Conclusions
In this study, we examined physicians' perceptions of the accuracy and comprehensibility of ChatGPT's responses to IVF FAQs and their attitudes toward ChatGPT as a tool for patient education. When blinded to the source of responses, most physicians at least agreed that ChatGPT’s responses were accurate and comprehensible. However, when asked specifically about ChatGPT, physicians had more cautious attitudes regarding the reliability of ChatGPT for medical information.

Support
None

References
1. Mnajjed L, Patel RJ. Assessment of ChatGPT generated educational material for head and neck surgery counseling. Am J Otolaryngol. 2024 Jul 21;45(5):104410. doi: 10.1016/j.amjoto.2024.104410. Epub ahead of print. PMID: 39059168.
2. Momenaei B, Wakabayashi T, Shahlaee A, Durrani AF, Pandit SA, Wang K, Mansour HA, Abishek RM, Xu D, Sridhar J, Yonekawa Y, Kuriyan AE. Appropriateness and Readability of ChatGPT-4-Generated Responses for Surgical Treatment of Retinal Diseases. Ophthalmol Retina. 2023 Oct;7(10):862-868. doi: 10.1016/j.oret.2023.05.022. Epub 2023 Jun 3. PMID: 37277096.
3. Patel JM, Hermann CE, Growdon WB, Aviki E, Stasenko M. ChatGPT accurately performs genetic counseling for gynecologic cancers. Gynecol Oncol. 2024 Apr;183:115-119. doi: 10.1016/j.ygyno.2024.04.006. Epub 2024 Apr 26. PMID: 38676973.