

COMPARATIVE ANALYSIS OF CHAT-GPT RESPONSES OF IVF-RELATED QUESTIONS AND ASRM/ACOG ANSWERS



Morgan Wiley, MD; Kristina Hrvojevic, MD; Arlene Go, MD

Morgan.wiley@aah.org

Advocate Health; Endeavor Health; Fertility Center of Illinois

Background

- Use of AI tools such as ChatGPT has rapidly expanded in healthcare, with patients increasingly using these platforms to obtain medical information.
- Patients undergoing IVF often have questions about treatment protocols, risks, and expected outcomes.
- While AI tools may improve access to information, inaccurate or incomplete responses may pose risks to patient understanding and clinical counseling.

Objective

- To evaluate the accuracy and readability of ChatGPT responses to common IVF-related patient questions compared with recommendations from ASRM and ACOG.

Methods

Study Design

- Descriptive evaluation study comparing ChatGPT (GPT-5) responses to established clinical guidance from ASRM and ACOG.

Sample and Setting

- Twenty-two patient-style IVF questions were developed using online fertility forums and patient education materials.
- All questions were entered into ChatGPT (GPT-5) in a standardized manner.

Procedures

The questions were recorded verbatim and independently scored for accuracy against ASRM/ACOG guidance using a 0–2 scale adapted from Shirefraw et al. (2024). Inter-reviewer agreement and readability were assessed using Cohen's kappa and the Flesch–Kincaid Grade Level, respectively.

Methods (continued)

Measures / Instruments

- Accuracy of ChatGPT (GPT-5) responses was evaluated using a 3-point ordinal scale adapted from Shirefraw et al. (2024):

• 0 = incorrect, 1 = partially correct, 2 = correct

- Responses were assessed against current guidance from the ASRM and/or ACOG.

- Readability was measured using the Flesch–Kincaid Grade Level, a validated instrument for assessing reading difficulty.

Data Analysis

- Mean accuracy scores and readability grade levels were calculated for ChatGPT responses.

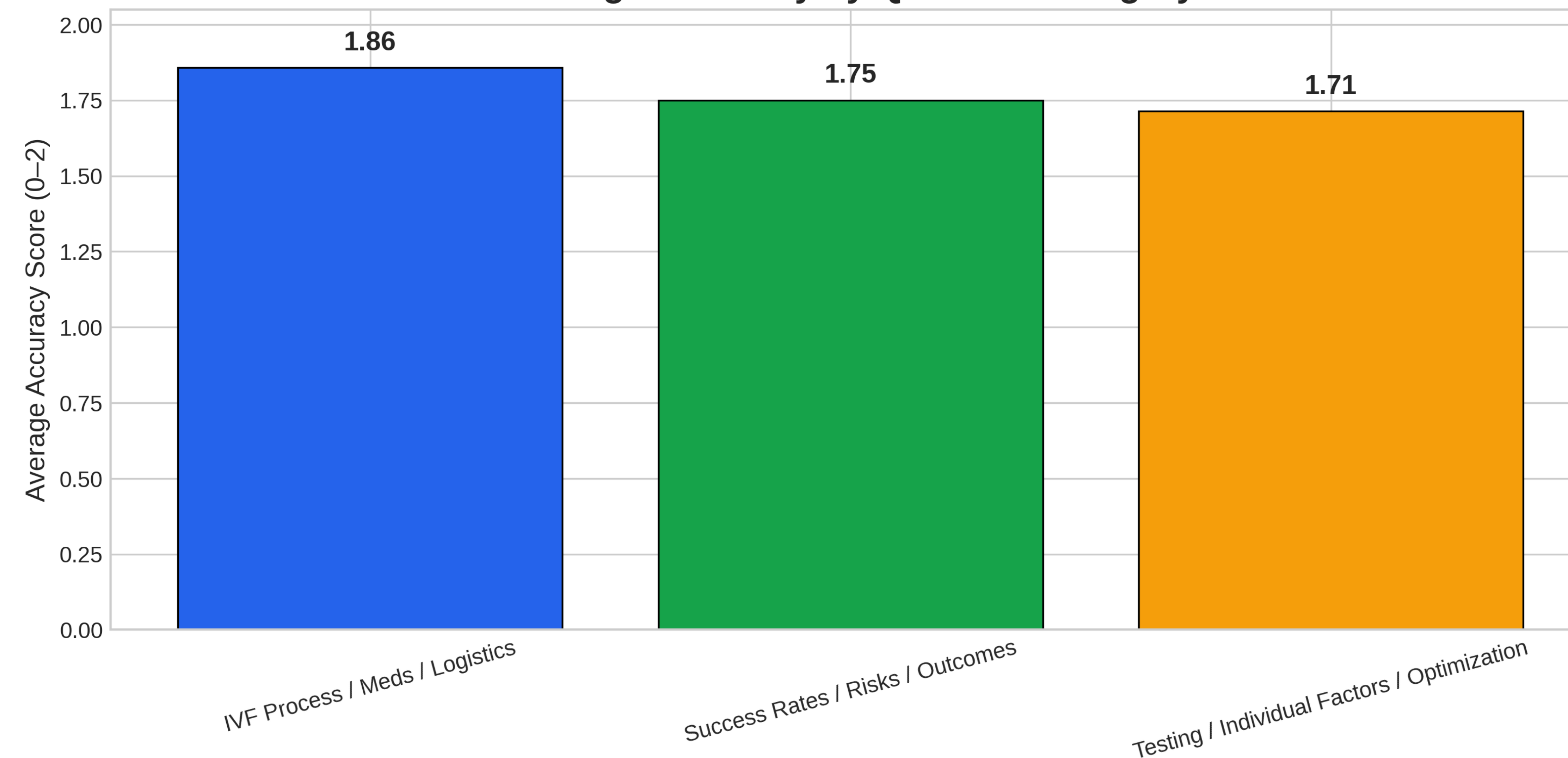
Results

- A total of **22 patient-style IVF questions** were evaluated.
- The **mean accuracy score** across all ChatGPT responses was **1.72 out of 2**, indicating frequent alignment with ASRM and/or ACOG guidance, though incomplete concordance was observed.
- **Inter-rater reliability** between reviewers demonstrated **moderate agreement** (Cohen's $\kappa = 0.57$).
- The **mean Flesch–Kincaid Grade Level** was **10.2**, suggesting that responses may be difficult for patients with lower health literacy to fully comprehend.

Sample Questions

- 1. I'm considering IVF — can you explain the basic steps involved?
- 4. How many embryos should usually be transferred during IVF?
- 10. What is ovarian hyperstimulation syndrome and how risky is it?
- 15. How are the IVF shots given, and are they painful?
- 17. What is PGT-A testing of embryos, and how does it work?

Average Accuracy by Question Category



Conclusions

- ChatGPT responses generally aligned with ASRM and ACOG guidance but were often incomplete, suggesting potential value as a supplemental information source rather than a standalone counseling tool.
- These findings are consistent with prior studies showing variable accuracy of AI for nuanced clinical questions, particularly in specialty care.
- Readability levels exceeded recommended standards for patient education, which may limit usefulness for patients with lower health literacy. However, AI users can request rephrasing if needed.
- **Limitations:** This study evaluated a limited number of questions from a single AI model at one time point and did not assess patient interpretation or real-world clinical impact.

Implications for Practice

These findings apply to clinical practice, as healthcare professionals often address AI-informed patient questions during counseling and education.

Future Research:

Future studies should assess additional AI platforms, broader question sets, and the impact of AI-generated information on patient understanding and clinical decision-making.

References

- **American Society for Reproductive Medicine (ASRM).** Practice Committee documents and patient education materials on assisted reproductive technology and IVF.
- **American College of Obstetricians and Gynecologists (ACOG).** Clinical guidance and patient education resources related to infertility and assisted reproductive technology.
- **Shirefraw A, et al.** Evaluation of ChatGPT responses to healthcare-related questions using a structured accuracy scoring system. 2024.
- **Flesch R, Kincaid JP.** A readability formula for evaluating written patient education materials.