

Expert-AI Concordance in Varicocele Management: How Reliable Is ChatGPT-4.0?

✉ Fahri Yavuz İlki¹, ✉ Emre Bülbül², ✉ Yusuf Kadir Topçu¹, ✉ Selahattin Bedir¹

¹University of Health Sciences Türkiye, Gülhane Training and Research Hospital, Clinic of Urology, Ankara, Türkiye

²Vakıfkebir State Hospital, Trabzon, Türkiye

What's known on the subject? and What does the study add?

Large language models like ChatGPT are increasingly used by both clinicians and patients to obtain medical information. However, their accuracy and alignment with expert recommendations in specific urological conditions such as varicocele remain unclear. Previous studies evaluating ChatGPT have focused mainly on general medical knowledge or patient education. This study is the first to systematically assess the clinical reliability of ChatGPT-4.0 in varicocele-related infertility scenarios using structured expert evaluation. It demonstrates that ChatGPT-4.0 provides recommendations largely consistent with expert opinion, especially in guideline-based standard cases, while highlighting its limitations in gray-zone decisions that require individualized clinical judgment.

Abstract

Objective: Artificial intelligence (AI)-based large language models (LLMs), such as ChatGPT-4.0, are increasingly being considered for clinical decision-making support. However, their reliability in providing clinical recommendations for varicocele-related infertility remains to be thoroughly evaluated. This study aimed to evaluate the reliability of ChatGPT-4.0 in providing clinical recommendations for patients with varicocele-related infertility.

Materials and Methods: A standardized clinical scenario was created involving a 32-year-old male with varicocele and oligoasthenoteratozoospermia, including details from physical examination, hormonal profile, and semen analysis based on the World Health Organization 6th edition criteria. Sixteen diagnostic and therapeutic questions were developed and submitted to ChatGPT-4.0. The AI-generated responses were reviewed by 24 experienced urologists specializing in varicocele management, who rated the recommendations using a 5-point Likert scale.

Results: The urologists demonstrated an 80.2% agreement, 10.7% disagreement, and 9.1% neutrality with ChatGPT-4.0 recommendations. For 14 of the 16 questions, the majority of urologists either agreed or strongly agreed with ChatGPT-4.0. Recommendations regarding varicocelectomy indication, antioxidant usage, the female partner age greater than 35, follow-up after varicocelectomy, testosterone deficiency, and normospermic varicocele showed the highest consensus. However, lower agreement rates were noted for microsurgical varicocelectomy (54.1%) and preoperative sperm cryopreservation (16.7%).

Conclusion: ChatGPT-4.0 demonstrates reliability in providing clinical recommendations in most scenarios related to varicocele treatment, showing strong agreement with expert clinicians. However, specific "gray zone" scenarios requiring individualized decision-making highlight limitations; emphasizing the importance of experienced clinical judgment. ChatGPT-4.0 can serve as a reliable informational tool regarding varicocele treatment but should be used with caution in complex clinical decisions requiring personalized evaluation.

Keywords: Varicocele, male infertility, artificial intelligence, ChatGPT-4.0

Correspondence: Fahri Yavuz İlki MD, University of Health Sciences Türkiye, Gülhane Training and Research Hospital, Clinic of Urology, Ankara, Türkiye

E-mail: fahri.ilki@istanbul.edu.tr **ORCID-ID:** orcid.org/0000-0001-7067-3815

Received: 30.07.2025 **Accepted:** 19.08.2025 **Epub:** 21.08.2025 **Publication Date:** 02.03.2026

Cite this article as: İlki FY, Bülbül E, Topçu YK, Bedir S. Expert-AI concordance in varicocele management: how reliable is ChatGPT-4.0? J Urol Surg. 2026;13(1):39-45.

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of the Society of Urological Surgery.

This is an open access article under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.



Introduction

Male infertility is a significant reproductive health issue affecting approximately 15% of couples, with male factors being the primary cause in 40–50% of these cases (1). One of the most common causes of male infertility is varicocele, which is observed in 35–40% of subfertile men and up to 15% of the general population (2,3). The effects of clinical varicocele on infertility have been studied for many years, and surgical treatment—particularly microscopic subinguinal varicolectomy—has been reported to improve sperm parameters and increase spontaneous pregnancy rates (4,5).

The management of varicocele cases remains controversial in "gray zone" situations such as mild sperm abnormalities, grade I varicocele, or azoospermia (6,7). Treatment decisions are typically based on a holistic evaluation of multiple factors, including the patient's clinical findings, semen analysis results, female partner's age, and the couple's reproductive expectations (8,9).

In recent years, the integration of artificial intelligence (AI) systems into clinical decision-making processes has accelerated. Artificial intelligence-based large language models (LLMs), especially advanced versions such as ChatGPT-4.0, have begun to be used experimentally in medical education and clinical decision support (10). Nowadays, patients frequently consult internet sources and LLMs for health-related issues. However, the extent to which LLMs are beneficial to patients and the accuracy of their recommendations in specific clinical scenarios remain unclear (11). The consistency of LLMs with expert opinions and the reliability of the information they provide have been assessed in only a limited number of studies (12).

In this study, diagnostic and therapeutic suggestions provided by ChatGPT-4.0 were obtained for an infertile male patient diagnosed with varicocele, as well as, for different clinical scenarios of varicocele. The clinical responses generated by ChatGPT for each scenario were evaluated by urology specialists. The aim was to evaluate the reliability of artificial intelligence-based language model recommendations by clinicians for varicocele disease.

Materials and Methods

Study Design

No patient identifiers were used in this study. LLMs are based on publicly available information. Since there were no human subjects involved, ethical approval was not required for studies related to LLMs (13,14). In this study, an AI-based language model, ChatGPT-4.0, was utilized. This observational study was conducted in April and May 2025. The Institutional

Ethics Committee of University of Health Sciences Türkiye, Gülhane Training and Research Hospital reviewed the study design and concluded that ethical approval was not necessary. This decision was based on the absence of patient data, the exclusive collection of anonymous professional opinions, and the non-interventional nature of the survey. Real patient data was not included in this study. No identifiable personal or health information was obtained, and all responses remained anonymous. The study was conducted in accordance with the principles outlined in the Declaration of Helsinki (2004 revision).

Creating Case Scenarios

The survey questions were created by the authors (F.Y.I., E.B., Y.K.T. and S.B.). A case scenario was designed based on a 32-year-old infertile male patient diagnosed with varicocele. The case included physical examination findings, hormone profile, scrotal Doppler ultrasonography results, and semen analysis (according to the 6th edition of the World Health Organization classification) (15). Subsequently, multiple clinical scenarios were generated by modifying various parameters such as sperm count, female partner's age, and hormone levels. A total of 16 questions were developed.

The initial scenario used in this study was based on a representative index case constructed to reflect a typical clinical presentation of varicocele-related infertility. To evaluate a broader range of decision-making contexts, this base scenario was systematically modified to include varying clinical parameters, thereby creating more complex clinical situations. Although no real patient data were used, the final set of scenarios was designed to closely resemble real-life cases.

To assess the clinical realism of the case scenarios, participants were asked to rate the similarity of the scenarios to actual clinical practice using a 5-point Likert scale. Remarkably, 100% of the respondents selected "strongly agree", indicating that the scenarios were perceived as highly reflective of real-world cases.

AI-based Clinical Decision Generation and Expert Opinion Survey

Diagnostic and therapeutic recommendations for infertile male case scenarios with varicocele were generated by ChatGPT-4.0. These responses were evaluated by 24 specialists in a survey of urologists. The evaluation was conducted and recorded through Google Forms. Urologists with experience in the diagnosis, follow-up, and treatment of varicocele were included in the study. Participants assessed the responses of ChatGPT-4.0 anonymously. Participation was voluntary, and no incentives were provided. No random sampling was performed. The online survey utilized a 5-point Likert scale to evaluate ChatGPT-4.0. The specialists were asked to rate their level of agreement with each recommendation using the Likert scale (1 = strongly disagree, 5 = strongly agree).

Statistical Analysis

The collected data were analyzed statistically using SPSS version 26.0 (IBM, Armonk, NY, USA). The data distribution was tested using the Kolmogorov–Smirnov test. For the comparison based on the experience level of urologists, an independent-samples t-test was used to assess the mean Likert scale scores. A p-value of <0.05 was considered statistically significant. Inter-rater agreement among the 24 participating urologists was assessed using Fleiss' kappa for multiple raters. The original 5-point Likert scale responses were dichotomized into "Agree" (scores 4-5) and "Not Agree" (scores 1-3) to facilitate interpretation. Kappa values were interpreted according to Landis and Koch's criteria, where <0.20 indicates slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and >0.80 almost perfect agreement.

Results

Clinical Features

A 32-year-old male patient presented with a two-year history of infertility. His 29-year-old female partner had no abnormalities on gynecological evaluation. The patient's testicular volumes were 18 mL on the right and 16 mL on the left. Physical examination revealed a grade II varicocele on the left side. Scrotal Doppler ultrasonography showed a left pampiniform plexus vein diameter of 3.4 mm during the Valsalva maneuver, with a reflux duration of 2.1 seconds. Hormonal analysis revealed follicle-stimulating hormone levels were 4.1 IU/L, luteinizing hormone levels were 3.5 IU/L, and total testosterone levels were 4.8 ng/mL. The semen analysis results of the representative case created for the study are presented in Table 1, which were consistent with varicocele-associated oligoasthenoteratozoospermia.

ChatGPT-4.0 Responses and Survey Evaluation

Among the 24 participating urology specialists, 12 had <5 years (group A) and 12 had >5 years (group B) of clinical experience. Additionally, twelve were employed at tertiary care centers, eight at state hospitals, and four in private institutions. The clinical questions posed to ChatGPT-4.0 and its corresponding responses are presented in Table 2. There were no statistically significant differences between groups A and B in terms of the mean scores assigned to any of the questions (p<0.05).

The distribution of responses from 24 urologists to the 16 clinical questions is illustrated in Figure 1. Each stacked bar depicts the proportion of participants selecting each category on the 5-point Likert scale, ranging from "Strongly disagree" to "Strongly agree." Visual inspection of the figure shows clear patterns of consensus in certain scenarios, particularly those with guideline-based management recommendations, while other scenarios, especially those in "gray zones," reveal more heterogeneous response distributions. This graphical representation complements the tabulated data and highlights the variation in expert agreement across all questions. Urologists who evaluated ChatGPT-4.0's response to 16 questions had 80.2% agreement, 10.7% disagreement, and 9.1% neutral responses. The distribution of expert ratings for each question is shown in Figure 1. For 14 out of 16 questions, the vast majority of urologists selected either "Agree" or "Strongly Agree" with ChatGPT-4.0.

The ChatGPT-4.0 responses regarding varicocelectomy indication (Question 1, 100%), antioxidant recommendation (Question 3, 91.7%), female partner age >35 (Question 6, 91.7%), follow-up after varicocelectomy (Question 7, 91.7%), testosterone deficiency (Question 10, 91.7%), and normospermic varicocele (Question 12, 95.8%) had the highest consensus among clinicians.

Table 1. Semen analysis results of the sample case presented to ChatGPT-4.0 and urology specialists

Parameter	1 st semen analysis	2 nd semen analysis
Ejaculate volume	2.6 mL	2.8 mL
Sperm concentration	9 million/mL	13 million/mL
Total sperm count	23.4 million	36.4 million
Progressive motility (PR)	18%	22%
Total motility (PR + NP)	32%	38%
Immotile sperm	68%	62%
Normal morphology (kruger)	3%	2%
Vitality (eosin test)	68%	70%
Presence of leukocytes	<1 million/mL (normal)	<1 million/mL (normal)
Agglutination	Not observed	Not observed
pH	7.8	7.7
Viscosity	Normal	Normal
Liquefaction time	20 minutes	20 minutes
NP: Non-progressive motility		

Table 2. A compilation of 16 clinical questions related to the diagnosis and management of varicocele, along with the corresponding answers generated by ChatGPT-4.0 in the context of a standardized male infertility case

Questions	Answers
Q1: Is varicocelectomy surgery indicated in this patient?	A1: This patient has symptomatic grade II varicocele, impaired semen parameters, and meets the criteria for surgical indication in the context of an infertile couple. Microsurgical subinguinal varicocelectomy is recommended.
Q2: In addition to varicocele surgery, would you consider a supportive approach in terms of medical treatment?	A2: In this patient, I would recommend antioxidant supplementation during both the preoperative and postoperative periods of varicocelectomy. Although the current hormonal profile does not require intervention, fertility-friendly agents such as clomiphene may be considered if a decline is observed during follow-up. Lifestyle modifications should also be strongly encouraged.
Q3: What types of preparations would you recommend as antioxidant therapy?	A3: In male patients with OAT for whom varicocelectomy is planned, a combination of antioxidants (e.g., CoQ10, L-carnitine, zinc, and vitamin E) may be recommended for a duration of 3 to 6 months. This supportive therapy has the potential to enhance sperm quality when used in conjunction with surgery.
Q4: If we go back to the beginning of the case, you mentioned that you would recommend varicocelectomy. If microsurgical equipment is not available, would you refer the patient to another center, or would you proceed with the surgery yourself?	A4: If I do not have the appropriate equipment or experience for microsurgical varicocelectomy, I would prefer to refer the patient to a center where this technique can be performed.
Q5: Would you recommend sperm cryopreservation for this patient prior to varicocelectomy?	A5: Although sperm cryopreservation is not an absolute indication in this case, the patient should be informed about the potential risk of deterioration in semen parameters following surgery, and the option of 'preventive sperm banking' should be offered.
Q6: If the female partner were 36 years old in this case, would it change your approach?	A16: When the female partner is 36 years old, the indication for varicocelectomy does not change, but the management strategy does: sperm cryopreservation is strongly recommended to avoid loss of time, a parallel plan with IVF is considered, and the observation period is shortened.
Q7: Assuming varicocelectomy has been performed in this patient, how long should we wait for spontaneous pregnancy postoperatively?	A7: If there is no urgent reproductive pressure for the couple after varicocelectomy, I would wait at least 3 months, preferably 6 months, for spontaneous conception. If pregnancy does not occur within this period, I would reassess both the sperm parameters and the condition of the female partner to revise the treatment plan accordingly.
Q8: If the patient has already cryopreserved sperm prior to varicocelectomy, would you still wait for 6 months or consider extending the waiting period?	A8: If sperm has been cryopreserved, the waiting period for natural conception can be extended. However, depending on the couple's specific circumstances, this period should not exceed 12 months.
Q9: If the first semen analysis showed OAT, but the second one was within normal limits, what would you recommend?	A9: The female partner's age of 29 places this case at the lower threshold for surgical indication. Due to the variability in semen analysis, a third test should be requested; if stable normozoospermia is observed, there would no longer be an indication for surgery. Since the chance of natural conception remains high, active surveillance is the most appropriate strategy.
Q10: If this patient had mild testosterone deficiency, how would it affect your treatment plan?	A9: In infertile men with varicocele and mild testosterone deficiency, varicocelectomy should be prioritized, and the hormonal response should be monitored postoperatively. If the testosterone deficiency is symptomatic, fertility-preserving treatments such as clomiphene or hCG may be considered. Direct testosterone replacement is contraindicated in the context of fertility.
Q11: If this patient had a left-sided grade 1 varicocele, what would your treatment plan be?	A10: In an infertile male with grade I varicocele and mild semen abnormalities, active surveillance is preferred prior to surgery, especially if the female partner is young and there is no time pressure. Surgery is considered only if semen parameters worsen or if the couple opts against assisted reproductive techniques such as IVF.
Q12: If both semen analyses had shown normal sperm parameters in this case, would you still recommend varicocelectomy?	A11: No, if both semen analyses are normal and the patient is asymptomatic, I would not recommend varicocelectomy. Instead of surgery, monitoring the chances of natural conception and adopting a couple-based approach to infertility would be preferred.

Table 2. Continued	
Questions	Answers
Q13: If the semen analysis results in this case had shown severe OAT, what would your initial recommendation be?	A12: If the semen analysis findings are severely impaired in this case, sperm cryopreservation should be performed prior to surgery. Microsurgical varicocelectomy should then be planned. Although an appropriate female partner age may allow for a period of observation, the couple should be counseled in advance about the possibility of treatment failure.
Q14: If azoospermia had been detected in the semen analysis of this patient, what would your initial management plan be?	A13: Initially, genetic testing and hormonal evaluation should be performed. Depending on the patient's condition, micro-TESE, varicocelectomy, or a combination of both may be planned. However, the decision for surgery should be based on the potential for sperm production.
Q15: Which should be performed first: micro-TESE or varicocelectomy?	A14: If the patient has non-obstructive azoospermia and is a suitable surgical candidate (e.g., AZFc deletion, FSH <15, and moderate testicular volume), varicocelectomy may be recommended as the initial approach. This is because some patients may exhibit spontaneous sperm in the ejaculate postoperatively, or the success rate of subsequent micro-TESE may improve.
Q16: Would you consider performing micro-TESE and varicocelectomy simultaneously?	A15: In azoospermic patients with limited testicular reserve, time constraints, and a planned IVF cycle, performing micro-TESE and varicocelectomy in the same session may be considered. However, this approach requires careful surgical planning and informed patient consent. In most cases, performing varicocelectomy first and observing the outcome is a safer and more physiological strategy.
OAT: Oligoasthenoteratozoospermia, TESE: Testicular sperm extraction, FSH: Follicle stimulating hormone, CoQ: Coenzyme Q, IVF: In vitro fertilization, hCG: Human chorionic gonadotropin, AZFc: Azoospermia factor c	

The discussion between ChatGPT-4.0 and urologists regarding microsurgical varicocelectomy (Question 4) showed an agreement rate of 54.1% (13/24), a neutral rate of 20.9% (5/24), and a disagreement rate of 25% (6/24). Regarding preoperative sperm cryopreservation (Question 5), the agreement rate among ChatGPT-4.0 and urologists was 16.7% (4/24), the neutral rate was 33.3% (8/24), and the disagreement rate was 50% (12/24).

Inter-rater agreement was re-evaluated by considering only the "Agree" (scores 4-5) and "Disagree" (scores 1-2) categories, excluding "Neutral" (score 3) responses. Using this approach, the overall Fleiss kappa value was 0.267, indicating fair to moderate agreement according to the Landis and Koch classification. This method eliminates the influence of indecisive responses and allows a clearer assessment of the consistency between positive and negative judgments.

Discussion

This study assessed the concordance between the clinical recommendations of ChatGPT-4.0 for varicocele-related infertility and the evaluations of practicing urologists. The high overall agreement suggests that ChatGPT-4.0 is capable of generating recommendations aligned with expert perspectives, particularly when guideline-based indications are clear.

Disagreement was most evident in scenarios involving microsurgical varicocelectomy and preoperative sperm

cryopreservation. These differences likely reflect variations in institutional resources, surgical expertise, and the absence of definitive guidance in current clinical recommendations. For example, while microsurgical varicocelectomy is widely regarded as the standard of care, practical limitations may influence decision-making in certain settings. Similarly, the lack of strong evidence regarding preoperative sperm cryopreservation leads to heterogeneity in clinical practice.

The model's suggestion to incorporate antioxidant therapy and to adjust treatment strategies when the female partner is older than 35 years was broadly endorsed by experts, reflecting alignment with both current literature (4,5) and international guidelines (8,9). The consistency of agreement across urologists with different levels of experience further suggests that LLMs may offer uniform, guideline-consistent information to patients, regardless of the clinician's background.

Despite these strengths, limitations remain in "gray zone" scenarios, where nuanced clinical judgment is essential and standardized recommendations are lacking. This reinforces the view expressed in prior literature (16,17) that AI tools should serve as complementary aids rather than replacements for clinician expertise.

Study Limitations

This study has several limitations. First, the performance of ChatGPT-4.0 was evaluated based on fixed clinical scenarios,

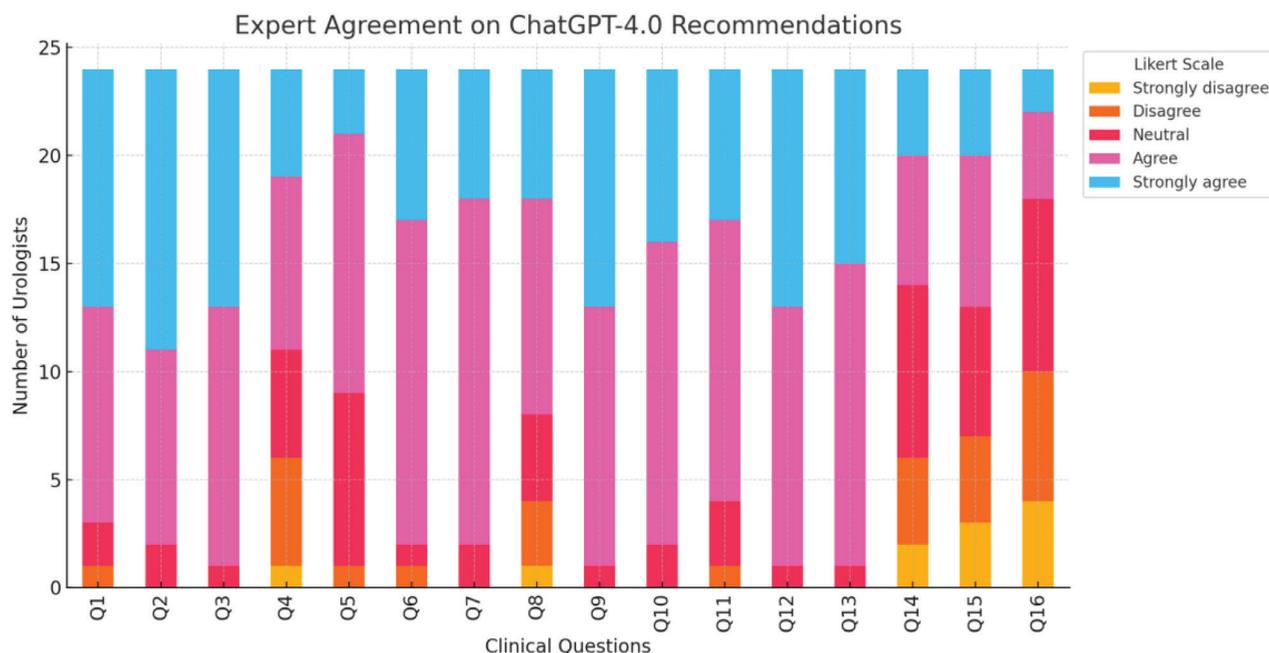


Figure 1. Distribution of responses from 24 urologists evaluating the appropriateness of ChatGPT-4.0's answers to 16 clinical questions using a 5-point Likert scale. Each stacked bar represents the number of experts selecting each level of agreement (from "Strongly disagree" to "Strongly agree") for a given question. The graphic illustrates the degree of expert consensus or divergence across all responses

which may not fully reflect the dynamic nature of real-life patient interactions. Second, the number of participating urologists was limited, which may restrict the generalizability of expert opinions. Previous studies have shown that LLMs may occasionally produce inaccurate responses that appear correct and may demonstrate inconsistencies when dealing with complex clinical scenarios (18–20). Therefore, although ChatGPT-4.0 showed high agreement with expert opinions in many scenarios, it should be used cautiously in complex cases.

Conclusion

While AI-based language models such as ChatGPT-4.0 demonstrate a high level of consistency with clinical guidelines and expert consensus in standard index cases, their utility remains limited in complex or low-evidence scenarios where nuanced clinical judgment is essential. In such contexts, expert opinion remains irreplaceable, and current AI systems cannot substitute for the depth and flexibility of experienced clinical reasoning.

Ethics

Ethics Committee Approval: Not necessary.

Informed Consent: Not necessary.

Authorship Contributions

Footnotes

Surgical and Medical Practices: Y.K.T., S.B., Concept: F.Y.İ., Design: F.Y.İ., Data Collection or Processing: Y.K.T., Analysis or Interpretation: E.B., Literature Search: E.B., Writing: F.Y.İ.

Conflict of Interest: No conflict of interest was declared by the authors.

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

References

1. Agarwal A, Baskaran S, Parekh N, Cho CL, Henkel R, Vij S, Arafa M, Panner Selvam MK, Shah R. Male infertility. *Lancet*. 2021;397:319–333. [\[Crossref\]](#)
2. Sharlip ID, Jarow JP, Belker AM, Lipshultz LI, Sigman M, Thomas AJ, Schlegel PN, Howards SS, Nehra A, Damewood MD, Overstreet JW, Sadovsky R. Best practice policies for male infertility. *Fertil Steril*. 2002;77:873–882. [\[Crossref\]](#)
3. Baazeem A, Belzile E, Ciampi A, Dohle G, Jarvi K, Salonia A, Weidner W, Zini A. Varicocele and male factor infertility treatment: a new metaanalysis and review of the role of varicocele repair. *European Urology*. 2011;60:796–808. [\[Crossref\]](#)
4. Esteves SC, Miyaoka R, Roque M, Agarwal A. Microsurgical varicocelectomy improves fertility outcomes. *Asian J Androl*. 2016;18:246–255. [\[Crossref\]](#)
5. Wang R, Zhang Z, Zhang Y, Wang Y, Wang J, Zhang Q, Wang C. Efficacy of varicocelectomy on semen parameters and pregnancy rate. *Fertil Steril*. 2018;110:897–905.e1 [\[Crossref\]](#)

6. Schlegel PN, Sigman M, Collura B, De Jonge CJ, Eisenberg ML, Lamb DJ, Mulhall JP, Niederberger C, Sandlow JI, Sokol RZ, Spandorfer SD, Tanrikut C, Treadwell JR, Oristaglio JT, Zini A. Diagnosis and treatment of infertility in men: AUA/ASRM guideline part II. *Fertil Steril*. 2021;115:62-69. [\[Crossref\]](#)
7. Esteves SC, Miyaoka R, Agarwal A. An update on the clinical assessment of the infertile male. [corrected]. *Clinics (Sao Paulo)*. 2011;66:691-700. Erratum in: *Clinics (Sao Paulo)*. 2012;67:203. [\[Crossref\]](#)
8. Minhas S, Boeri L, Capogrosso P, Cocci A, Corona G, Dinkelman-Smit M, Falcone M, Jensen CF, Gül M, Kalkanlı A, Kadioğlu A, Martinez-Salamanca JI, Morgado LA, Russo GI, Serefoğlu EC, Verze P, Salonia A. European Association of Urology Guidelines on male sexual and reproductive health: 2025 update on male infertility. *Eur Urol*. 2025;87:601-616. [\[Crossref\]](#)
9. AUA/ASRM guideline: the evaluation and treatment of male infertility. Available link: <https://www.asrm.org/practice-guidance/practice-committee-documents/diagnosis-and-treatment-of-infertility-in-men-uaa-asrm-guideline-part2> [\[Crossref\]](#)
10. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, Madriaga M, Aggabao R, Diaz-Candido G, Maningo J, Tseng V. Performance of ChatGPT on the United States medical licensing examination. *PLoS Digit Health*. 2023;2:e0000198. [\[Crossref\]](#)
11. Şahin MF, Keleş A, Özcan R, Doğan Ç, Topkaç EC, Akgül M, Yazıcı CM. Evaluation of information accuracy and clarity: ChatGPT responses to the most frequently asked questions about premature ejaculation. *Sex Med*. 2024;12:qfae036. [\[Crossref\]](#)
12. Stenzl A, Armstrong AJ, Rogers E, Habr D, Walz J, Gleave M, Sboner A, Ghith J, Serfass L, Schuler KW, Garas S, Chari D, Truman K, Sternberg CN. Evaluation of ChatGPT as a reliable source of medical information on prostate cancer for patients: global comparative survey of medical oncologists and urologists. *Urol Pract*. 2025;12:229-240. [\[Crossref\]](#)
13. Şahin MF, Topkaç EC, Doğan Ç, Şeramet S, Tuncer FB, Köroğlu MS, Orbeği O, Yazıcı CM. AI-assisted transcription of YouTube videos on penile enlargement: analysis of their text quality and readability. *Sex Med*. 2025;13:qfaf023. [\[Crossref\]](#)
14. Barlas İŞ, Tunç L. Quality of chatbot responses to the most popular questions regarding erectile dysfunction. *Urol Res Pract*. 2025;50:253-260. [\[Crossref\]](#)
15. World Health Organization. (2021). WHO laboratory manual for the examination and processing of human semen. In *Who laboratory manual for the examination and processing of human semen*. Available link: <https://www.who.int/publications/i/item/9789240030787> [\[Crossref\]](#)
16. Jeblick P, Schachtner B, Dextl J, Mittermeier A, Stüber AT, Topalis J, Weber T, Wesp P, Sabel B, Ricke J. ChatGPT makes mistakes in radiology. *Radiol Artif Intell*. 2023;5:e220302. [\[Crossref\]](#)
17. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, Chartash D. How well does ChatGPT answer medical questions? *J Med Internet Res*. 2023;25:e46599. [\[Crossref\]](#)
18. Patel SB, Lam K, Chatfield M, Furszyfer Del Rio D, Zou J, Szolovits P, Atreya S, Tufano J, Topol EJ. ChatGPT in healthcare: a taxonomy and systematic review. *NPJ Digit Med*. 2023;6:154. [\[Crossref\]](#)
19. Ben-Israel D, Jacobs R, White RW. Risks and limitations of AI models in clinical practice. *Lancet Digit Health*. 2023;5:e123-e125. [\[Crossref\]](#)
20. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, Moy L. ChatGPT and other large language models in healthcare: a review of opportunities and risks. *J Am Med Inform Assoc*. 2023;30:e230163. [\[Crossref\]](#)