E-ISSN: 2148-094X

istanbul MEDICAL JOURNAL

İSTANBUL TIP DERGİSİ

VOLUME 27 • ISSUE 2 • May 2026

Editorial

Original Articles

Letter to the Editor

istanbulmedicaljournal.org

galenos yayınevi

İstanbul MEDICAL JOURNAL

İSTANBUL TIP DERGİSİ

Publisher Contact
Address: Molla Gürani Mah. Kaçamak Sk. No: 21/1 34093 İstanbul, Türkiye
Phone: +90 (530) 177 30 97 / +90 (539) 307 32 03
E-mail: info@galenos.com.tr/yayin@galenos.com.tr
Web: www.galenos.com.tr
Publisher Certificate Number: 14521

Publication Date: May 2026
ISSN: 2619-9793 **E-ISSN:** 2148-094X
International periodical journal published four times in a year.

istanbulmedicaljournal.org

İSTANBUL MEDICAL JOURNAL
İSTANBUL TIP DERGİSİ

istanbul
MEDICAL JOURNAL
İSTANBUL TIP DERGİSİ

Pediatric Cardiovascular Surgery

Ali Can HATEMİ
Department of Pediatric Cardiovascular Surgery, University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital, İstanbul, Türkiye

Dermatology

Ayşe Esra KOKU AKSU
Department of Dermatology, University of Health Sciences Türkiye, İstanbul Training and Research Hospital, İstanbul, Türkiye

Mehmet Salih GÜREL
Department of Dermatology, İstanbul Medeniyet University Faculty of Medicine, İstanbul, Türkiye

Infectious Diseases and Clinical Microbiology

Nagehan Didem SARI
Department of Infectious Diseases and Clinical Microbiology, University of Health Sciences Türkiye, İstanbul Training and Research Hospital, İstanbul, Türkiye

Mustafa YILDIRIM
Department of Infectious Diseases and Clinical Microbiology, University of Health Sciences Türkiye, İstanbul Training and Research Hospital, İstanbul, Türkiye

Naim MAHROUM
Department of Infectious Diseases and Clinical Microbiology, Tel Aviv University Faculty of Medicine, Tel Aviv, İsrail

Physical Medicine and Rehabilitation

Ebru AYTEKİN
Department of Physical Medicine and Rehabilitation, University of Health Sciences Türkiye, İstanbul Training and Research Hospital, İstanbul, Türkiye

Nalan ÇAPAN
Department of Physical Therapy and Rehabilitation, İstanbul University, İstanbul Faculty of Medicine, İstanbul, Türkiye

General Surgery

Serkan SARI
Department of General Surgery, University of Health Sciences Türkiye, Başakşehir Çam and Sakura Training and Research Hospital, İstanbul, Türkiye

Mert Mahzuni SEVİNÇ
Department of General Medicine, University of Health Sciences Türkiye, İstanbul Training and Research Hospital, İstanbul, Türkiye

Ufuk Oğuz İDİZ
Department of General Medicine, University of Health Sciences Türkiye, İstanbul Training and Research Hospital, İstanbul, Türkiye

Hasan ÖKMEN
Department of General Medicine, University of Health Sciences Türkiye, İstanbul Training and Research Hospital, İstanbul, Türkiye

Chest Diseases

Erdoğan ÇETİNKAYA
Department of Chest Diseases, University of Health Sciences Türkiye, Yedikule Chest Diseases and Thoracic Surgery Training and Research Hospital, İstanbul, Türkiye

Seda TURAL ÖNÜR
Department of Chest Diseases, University of Health Sciences Türkiye, Yedikule Chest Diseases and Thoracic Surgery Training and Research Hospital, İstanbul, Türkiye

Sedat ALTIN
Department of Chest Diseases, University of Health Sciences Türkiye, Yedikule Chest Diseases and Chest Surgery Training and Research Hospital, İstanbul, Türkiye

Alexandru CORLATEANU
Department of Chest Diseases, State University of Medicine and Pharmacy, Moldova

Chest Surgery

Kemal KARAPINAR
Department of Chest Surgery, University of Health Sciences Türkiye, İstanbul Yedikule Pulmonary Diseases and Thoracic Surgery Training and Research Hospital, İstanbul, Türkiye

Levent CANSEVER
Department of Chest Surgery, University of Health Sciences Türkiye, Yedikule Chest Diseases and Chest Surgery Training and Research Hospital, İstanbul, Türkiye

Eye Diseases

Özen AYRANCI
Department of Eye Diseases, University of Health Sciences Türkiye, İstanbul Training and Research Hospital, İstanbul, Türkiye

Yusuf YILDIRIM
Department of Ophthalmology, University of Health Sciences Türkiye, İstanbul Beyoğlu Eye Training and Research Hospital, İstanbul, Türkiye

İSTANBUL MEDICAL JOURNAL
İSTANBUL TIP DERGİSİ

istanbul
MEDICAL JOURNAL
İSTANBUL TIP DERGİSİ

istanbul
MEDICAL JOURNAL
İSTANBUL TIP DERGİSİ

Please refer to the journal's webpage (**https://istanbulmedicaljournal.org/**) for "About Us" and "Instructions to Authors".

The İstanbul Medical Journal and/or its editors are members of ICMJE, COPE, WAME, CSE and EASE, and follow their recommendations.

The İstanbul Medical Journal is indexed in **Web of Science-Emerging Sources Citation Index, EBSCO, Gale, TUBITAK ULAKBIM TR Index, ProQuest, Embase, Türk Atif Dizini, Türk Medline, DOAJ, Ardi, Hinari, Goali, Oare, CAB International (CABI)** and **J-GATE**.

The journal is published online.

Owner: İstanbul Training and Research Hospital

Responsible Manager: Feray Akbaş

İSTANBUL MEDICAL JOURNAL
İSTANBUL TIP DERGİSİ

CONTENTS

İstanbul Med J 2026; 27(2): 95-6
DOI: 10.4274/imj.galenos.2026.280426

Back to the Future: From Bariatric Surgery to Pharmacotherapy in Obesity Management

Feray Akbaş

University of Health Sciences Türkiye, İstanbul Training and Research Hospital, Clinic of Internal Medicine, İstanbul, Türkiye

Obesity is a chronic, relapsing disease with a multifactorial etiology. It is associated with numerous comorbidities and complications, serving as a gateway to a broad spectrum of serious health conditions. Accordingly, therapeutic interventions targeting obesity may confer both prophylactic and disease-modifying effects across its associated comorbidity spectrum (1).

Behavioral modification remains the cornerstone of obesity management, as consistently emphasized by national and international guidelines (1-5). In addition to structured nutritional therapy and physical activity interventions, strategies targeting stress reduction and sleep optimization are also recommended (1). Furthermore, the incorporation of psychological support from the outset, tailored to all patients, is recognized as a critical component of comprehensive care (4).

In addition to behavioral modifications, pharmacotherapy and bariatric/metabolic surgery are implemented when clinically indicated. Historically, bariatric surgery has been regarded as the most effective intervention compared with pharmacological treatments and has typically been reserved as a later-line option when behavioral and medical therapies failed to achieve sufficient outcomes (6,7). However, with the recent introduction of novel obesity management medications (OMMs), this paradigm has shifted. Surgery is no longer uniformly considered the final or most effective step in obesity management, and pharmacotherapy has re-emerged as a central component in the treatment algorithm.

Although bariatric surgery appears to demonstrate superior long-term efficacy compared with OMMs, short-term outcomes with OMMs are increasingly reported to be comparable (8). Given the invasive nature of surgical interventions and their potential for serious adverse effects, it may be more appropriate to reserve bariatric surgery for patients at the upper extreme of stage 3 obesity, particularly those who have not responded adequately to other therapeutic options and who are already at a markedly elevated clinical risk.

Currently; orlistat, liraglutide, semaglutide and tirzepatide are available in Türkiye for the management of obesity (9). Orlistat, a gastric and pancreatic lipase inhibitor, is administered orally, whereas liraglutide and semaglutide—both GLP-1 receptor agonists—and tirzepatide, a dual GIP/GLP-1 receptor agonist, are administered as injectable therapies (10).

Beyond inducing substantial weight loss—up to approximately 25% with injectable agents (mean −20.2% for tirzepatide and −13.7% for semaglutide) (11)—these medications also provide clinically meaningful benefits for a range of obesity-related comorbidities. These include the prevention and treatment of prediabetes, metabolic syndrome, and type 2 diabetes, as well as broader cardiometabolic effects, such as reduction in major adverse cardiovascular events, improvement in chronic kidney disease, lowering of blood pressure, benefits in heart failure with preserved ejection fraction, and improvement in metabolic dysfunction–associated steatohepatitis. Additionally, they confer biomechanical benefits, including improvement in obstructive sleep apnea and osteoarthritis (12).

Thus, it would be fair to say that we are, in a sense, “back to the future” with respect to the use of pharmacotherapy for obesity. With the anticipated arrival of triple agonists and oral GLP-1–based agents, treatment strategies in obesity management are likely to undergo substantial transformation.

References

1. Busetto L, Dicker D, Frühbeck G, Halford JCG, Sbraccia P, Yumuk V, et al. A new framework for the diagnosis, staging and management of obesity in adults. Nat Med. 2024; 30: 2395-9.
2. Nadolsky K, Garvey WT, Agarwal M, Bonnecaze A, Burguera B, Chaplin MD, et al. American Association of Clinical Endocrinology Consensus Statement: algorithm for the evaluation and treatment of adults with obesity/adiposity-based chronic disease - 2025 update. Endocr Pract. 2025; 31: 1351-94.
3. Yumuk V, Tsigos C, Fried M, Schindler K, Busetto L, Micic D, et al. European guidelines for obesity management in adults. Obes Facts. 2015; 8: 402-24. Erratum in: Obes Facts. 2016; 9: 64.
4. Wharton S, Lau DCW, Vallis M, Sharma AM, Biertho L, Campbell-Scherer D, et al. Obesity in adults: a clinical practice guideline. CMAJ. 2020; 192: E875-91.

Address for Correspondence: Prof. Feray Akbaş, MD, University of Health Sciences Türkiye, İstanbul Training and Research Hospital, Clinic of Internal Medicine, İstanbul, Türkiye
E-mail: atlibatur@yahoo.com **ORCID ID:** orcid.org/0000-0001-5091-9160

Cite this article as: Akbaş F. Back to the future: from bariatric surgery to pharmacotherapy in obesity management. İstanbul Med J. 2026; 27(2): 95-6

5. Sonmez A, Sonmez A, Bayram F, Banli O, Demirci I, Kiyici S, et al. Türkiye Consensus Report on the multidisciplinary obesity treatment in adults. Obes Facts. 2025; 18: 514-34.

6. Gloy VL, Briel M, Bhatt DL, Kashyap SR, Schauer PR, Mingrone G, et al. Bariatric surgery versus non-surgical treatment for obesity: a systematic review and meta-analysis of randomised controlled trials. BMJ. 2013; 347: f5934.

7. Cui B, He J, Yao H. Effects of metabolic bariatric surgery on weight loss and diabetes remission over 10 years: a network meta-analysis of randomized controlled trials. Obes Surg. 2026; 36: 1786-96.

8. De Luca M, Cohen RV, Belluzzi A, Navarra G, Di Lorenzo N, Petry TBZ, et al. Efficacy and safety of pharmacological, endoscopic, and surgical treatments for obesity: a GRADE-based network meta-analysis. Obesity (Silver Spring). 2026; 34: 279-93.

9. Çetinkaya Altuntaş S. Medical treatment in obesity. Turk J Diab Obes. 2023; 7: 263-72.

10. McGowan B, Ciudin A, Baker JL, Busetto L, Dicker D, Frühbeck G, et al. A systematic review and meta-analysis of the efficacy and safety of pharmacological treatments for obesity in adults. Nat Med. 2025; 31: 3317-29.

11. Aronne LJ, Horn DB, le Roux CW, Ho W, Falcon BL, Gomez Valderas E, et al. Tirzepatide as compared with semaglutide for the treatment of obesity. N Engl J Med. 2025; 393: 26-36.

12. Guglielmi V, Bettini S, Sbraccia P, Busetto L, Pellegrini M, Yumuk V, et al. Beyond weight loss: added benefits could guide the choice of anti-obesity medications. Curr Obes Rep. 2023; 12: 127-46.

İstanbul Med J 2026; 27(2): 97-102
DOI: 10.4274/imj.galenos.2026.04742

Performance Evaluation of Large Language Models in Emergency Medicine Specialty Examination Questions: A Cross-Sectional Study

Şebnem Zeynep Eke Kurt[1], Suphi Bahadırlı[2]

[1]University of Health Sciences Türkiye, Taksim Training and Research Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye
[2]Medipol Mega University Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye

ABSTRACT

Introduction: Large language models (LLMs) have recently shown strong potential in medical education, yet their performance compared with human learners in specialty-level examinations remains unclear. This study aimed to evaluate the performance of LLMs compared to human groups on a 50-question emergency medicine test from the Turkish Medical Specialty Examination.

Methods: A cross-sectional study was conducted at İstanbul Medipol University in 2024, involving 40 medical students and postgraduates and six LLMs (ChatGPT 4o, Claude Sonnet 3.5, Gemini Advanced, ChatGPT 4.0 Mini, Gemini Flash, Claude Haiku). Participants completed a 50-question test. Correct answers were analyzed using Welch's one-way analysis of variance (ANOVA), Levene's test for homogeneity of variances, and Games-Howell post-hoc tests.

Results: Claude Sonnet 3.5 achieved the highest mean correct answers (46.4±0.548), followed by ChatGPT 4o (44.6±1.14) and Gemini Advanced (43.6±1.67). Postgraduates with 5+ years of experience scored 43.5±3.03, while fifth-year medical students scored the lowest (29.1±3.73). Welch's ANOVA indicated significant group differences [$F(9, 20.8): 31.3$, $p<0.001$]. Post-hoc tests revealed LLMs outperformed most human groups, with Claude Sonnet 3.5 significantly surpassing Claude Haiku (mean difference: 9.6, $p=0.028$).

Conclusion: LLMs demonstrated superior performance compared to most human groups, indicating their potential as educational tools in emergency medicine.

Keywords: Large language models, emergency medicine, medical education

Introduction

Emergency medicine requires quick and precise decision-making and skills, both of which are rigorously evaluated by examinations such as the Turkish Medical Specialty Examination (TUS). The TUS assesses competency through complex, scenario-based questions that demand clinical reasoning and practical knowledge (1). Developments in artificial intelligence (AI), particularly large language models (LLMs), have introduced tools capable of addressing medical queries, thereby increasing interest in their performance on standardized tests (2). When trained on extensive datasets, LLMs produce responses that resemble expert knowledge, suggesting applications in medical education and clinical support (3,4). However, their ability to address specialty-specific, time-sensitive questions in emergency medicine remains largely unexamined (5).

Few studies have compared LLMs with human learners in specialty examinations. Studies on general medical licensing exams, such as the United States Medical Licensing Examination (USMLE), have shown that LLMs, such as ChatGPT, outperform medical students (6,7). Emergency medicine, however, presents unique challenges; it requires rapid synthesis of clinical information under pressure and tests both human and AI capabilities (8). The literature lacks comparisons among LLMs, medical students at different training levels, and experienced postgraduates in this field (9). This gap is significant, as LLMs can enhance training by offering accessible educational resources, but their effectiveness in high-pressure specialties requires validation (10).

Unlike prior studies primarily focused on USMLE-style examinations, this study evaluates LLM performance using the TUS, which reflects a different linguistic and curricular context. By directly comparing LLMs

Address for Correspondence: Şebnem Zeynep Eke Kurt, MD, University of Health Sciences Türkiye, Taksim Training and Research Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye
E-mail: sebnemzeynepeke@hotmail.com **ORCID ID:** orcid.org/0000-0003-0778-8884

Cite this article as: Eke Kurt ŞZ, Bahadırlı S. Performance evaluation of large language models in emergency medicine specialty examination questions: a cross-sectional study. İstanbul Med J. 2026; 27(2): 97-102

Received: 24.11.2025
Accepted: 07.02.2026
Publication Date: 12.05.2026

with emergency medicine trainees and experienced specialists, this study provides specialty-specific evidence that extends existing AI examination literature.

This study aims to evaluate the performance of LLMs compared with that of medical students and postgraduates on a 50-question emergency medicine test from the TUS, using a cross-sectional design.

Methods

This cross-sectional study was approved by the İstanbul Medipol University Non-Interventional Clinical Research Ethics Committee (decision number: 664, date: 04.07.2024). The research was conducted at a Medipol Mega University Hospital Emergency Medicine Department in 2024. All participants provided informed consent prior to participation in the study. No patient data were collected or used in this research.

The study included 40 human participants: 10 fifth-year medical students, 10 sixth-year medical students, 10 first-year postgraduates, and 10 postgraduates with 5 or more years of experience in emergency medicine. Postgraduate participants consisted of emergency medicine residents (PG1) and board-certified emergency medicine specialists with at least five years of clinical experience (PG5+).

The inclusion criterion was enrollment in or graduation from a medical program, including postgraduates specializing in emergency medicine. No exclusion criteria were applied, as all eligible participants completed the test. Six LLMs "ChatGPT 4o, Claude Sonnet 3.5, Gemini Advanced, ChatGPT 4o Mini, Gemini Flash, and Claude Haiku" were also tested, each of which was evaluated five times to account for response variability. The participants were selected consecutively from the university's medical program and its emergency medicine department to ensure a representative sample.

Large Language Model Configuration and Prompting

The following LLMs were evaluated: ChatGPT-4o and ChatGPT-4o Mini (OpenAI; accessed May 2024), Claude Sonnet 3.5 and Claude Haiku (Anthropic; accessed June 2024), and Gemini Advanced and Gemini Flash (Google DeepMind; accessed June 2024).

All models were accessed through their official web-based interfaces using the default inference settings. Temperature, top-p, and related sampling parameters were not manually adjusted to reflect typical real-world user conditions and enhance reproducibility. A standardized prompt was used for all LLMs. Models were instructed to select the single best answer from the provided multiple-choice options without providing explanations or other commentary.

No chain-of-thought reasoning or clinical justification was explicitly requested. Prompt used for all LLMs: "You are answering a multiple-choice emergency medicine examination question. Select the single best answer (A, B, C, D, or E). Do not provide explanations or additional text."

The primary outcome was the number of correct answers on a 50-question multiple-choice test derived from past TUS examinations that focused on emergency medicine topics. Potential confounders, such as test familiarity or LLM version updates, were minimized by standardizing test conditions and using the latest model versions available in 2024. The test was validated by faculty experts for relevance and difficulty, ensuring content validity. The human participants completed the test under proctored conditions in a controlled environment.

While human participants completed the test under time-limited examination conditions, LLMs were not subject to time constraints.

The 50-item test was derived from TUS questions previously administered between 2018 and 2023. Items covered the core domains of emergency medicine, including trauma, toxicology, cardiology, neurology, infectious diseases, and critical care. Content validity was assessed by two senior emergency medicine faculty members who independently reviewed each item for relevance, clarity, and curriculum alignment. Disagreements were resolved by consensus. No items were modified from their original wording.

Each correct answer was scored as 1 point, with a maximum score of 50. Two researchers independently verified LLM responses, achieving high interrater reliability (kappa=0.95). No specific training was provided for data collection, but the process was standardized to reduce variability.

Selection bias was minimized by including all eligible participants consecutively, whereas information bias was reduced through standardized question presentation and scoring protocols. The variability in the LLM responses was addressed by conducting five test iterations per model and averaging the results. For LLMs, the unit of analysis was the model itself rather than individual runs. Five repeated runs were performed to estimate response variability; model-level mean scores and standard deviations (SDs) were used in all primary analyses.

The sample size was calculated to detect a five-point mean difference in the number of correct answers between groups, with 80% power and an alpha of 0.05, requiring at least 10 participants in each human group and five iterations per LLM, based on prior studies of medical examination performance. The number of correct answers was treated as a continuous variable and summarized via means and SDs.

Statistical Analysis

Statistical analyses included descriptive statistics to report means and SDs. Normality was confirmed via Shapiro-Wilk tests. Owing to nonhomogeneous variances [Levene's test, F (9, 60): 2.5, p=0.017], Welch's one-way analysis of variance (ANOVA) was used to assess differences in correct answers across groups, followed by Games–Howell post-hoc tests to identify specific group differences. As a sensitivity analysis, results were re-evaluated using model-level mean scores without treating individual LLM runs as independent observations. This approach yielded consistent group-level conclusions. No missing data were observed, so no imputation was needed. Analyses were performed using SPSS version 27. A p value <0.05 was considered significant. The primary outcome was the number of correct answers on the 50-question test.

Results

Participants

Forty human participants (mean age 28.5±4.2 years; 60% male) and six LLMs were included in the study. No exclusions occurred, and no data were missing.

Descriptive Data

Figure 1 shows the mean number of correct answers and the corresponding SDs. Table 1 presents the descriptive performance characteristics of all human groups and LLMs, including measures of central tendency and dispersion, to provide an overall comparison prior to inferential analyses. Claude Sonnet 3.5 scored highest (46.400±0.548), followed by ChatGPT 4o (44.600±1.140), and Gemini Advanced (43.600±1.670). Among human participants, PG5+ physicians scored 43.5±3.03, whereas fifth-year medical students scored the lowest (29.1±3.73).

Main Results

Welch's ANOVA revealed significant group differences [F (9, 20.8): 31.3, p<0.001)] (Table 2).

Games-Howell post-hoc tests (Table 3) indicated that Claude Sonnet 3.5 outperformed both Claude Haiku (mean difference: 9.6; p=0.028) and most human groups (e.g., MS5; mean difference: 17.3; p<0.001).

Table 1. Descriptive performance characteristics across study groups

Group	n (participants/runs)	Mean score	SD	Min–max
MS5	10 participants	29.1	3.73	23–35
MS6	10 participants	33.4	4.86	25–41
PG1	10 participants	38.4	4.21	31–45
PG5+	10 participants	43.5	3.03	38–48
ChatGPT-4o	5 runs	44.6	1.14	43–46
Claude Sonnet 3.5	5 runs	46.4	0.55	46–47
Gemini Advanced	5 runs	43.6	1.67	41–45
ChatGPT-4o Mini	5 runs	38.0	6.80	29–46
Gemini Flash	5 runs	40.3	3.94	35–45
Claude Haiku	5 runs	36.8	4.12	32–43

For human groups, n represents the number of participants. For LLMs, n represents the number of repeated runs performed to estimate response variability. LLMs: Large language models, SD: Standard deviation, MS5: Medical student-phase 5, MS6: Medical student-phase 6, PG1: Postgraduate one year, PG5+: Postgraduate 5+ years, Min-max: Minimum-maximum

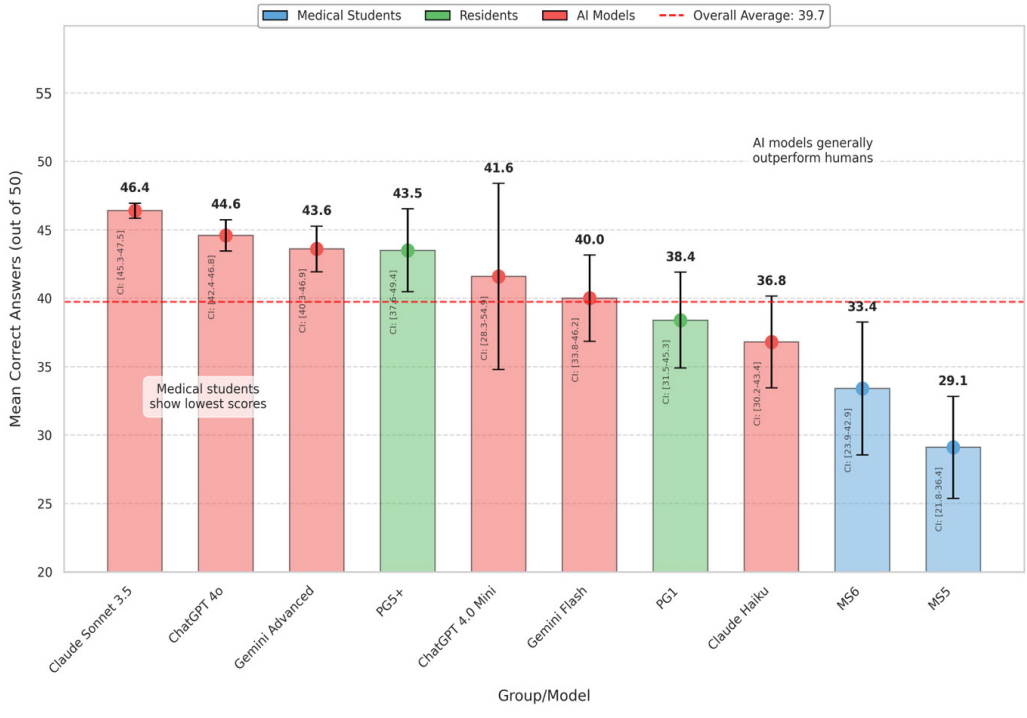

Figure 1. Demographics and analysis of responses to the test. AI: Artificial intelligence, MS5: Medical student-phase 5, MS6: Medical student-phase 6, PG1: Postgraduate one year, PG5+: Postgraduate 5+ years

The difference between Claude Sonnet 3.5 and fifth-year medical students was large (mean difference: 17.3 points; 95% confidence interval: 12.4–22.1; Cohen's d: 3.9). ChatGPT 4o and Gemini Advanced surpassed MS5 and MS6 ($p<0.001$). No significant differences were observed between PG5+ postgraduate physicians and top-performing LLMs (e.g., Claude Sonnet 3.5; mean difference: 2.9, $p=0.21$). While multiple pairwise comparisons were performed, the descriptive summary highlighted clear performance stratification across groups, which guided the interpretation of inferential results presented in Table 3.

Analysis of Performance Trends

The most proficient LLMs, Claude, Sonnet 3.5, ChatGPT 4o, and Gemini Advanced, exhibited noteworthy consistency, as indicated by their minimal SDs (0.548, 1.14, and 1.67, respectively). In contrast, ChatGPT 4o Mini exhibited the highest variability among all the groups (SD: 6.8), indicating potential inconsistencies in its performance. This higher variability may reflect architectural characteristics of lightweight models, which can be more sensitive to question phrasing and may lack domain-specific fine-tuning for emergency medicine content.

Among human groups, the performance gap between experience levels is notable. MS5 students, with the lowest mean score (29.1±3.73), were significantly outperformed by all LLMs and more experienced human groups ($p<0.001$ in most post-hoc comparisons).

This gap highlights the steep learning curve in emergency medicine, where foundational knowledge alone is insufficient without clinical exposure. MS6 students, after an additional year of training, improved to 33.4±4.86, but still lagged behind postgraduates and LLMs. The post-hoc analysis revealed that the difference between MS5 and MS6 was not statistically significant (mean difference: 4.3, $p=0.48$), suggesting that an additional year of medical school training may not substantially enhance performance on specialty-specific exams without targeted emergency medicine experience. Postgraduates with 5+ years of experience (PG5+) achieved a mean score of 43.5±3.03 and closely approached the performance of the top LLMs. The absence of a significant difference between PG5+ and Claude Sonnet 3.5 (mean difference: 2.9, $p=0.21$) or between PG5+ and ChatGPT 4o (mean difference: 1.1, $p=0.985$) underscores the competitive performance of experienced clinicians in this domain. However, their SD (3.03) indicates greater variability than that of the top LLMs, possibly reflecting differences in individual expertise or test-taking strategies among the participants.

Levene's test result [$F(9, 60)$: 2.5, $p=0.017$] indicates heterogeneity of variances across groups, consistent with the observed differences in SDs. For example, Claude Sonnet 3.5's exceptionally low SD (0.548) contrasts sharply with ChatGPT 4o Mini's high SD (6.8). This heterogeneity justified the use of Welch's ANOVA and Games–Howell post-hoc tests, which are robust to unequal variances.

The post-hoc analysis further reveals that the performance hierarchy among LLMs is not uniform. The significant outperformance of Claude Sonnet 3.5 relative to Claude Haiku (mean difference: 9.6, $p=0.028$) indicates substantial differences in capability even within the same family of models. Similarly, ChatGPT 4o outperforms Claude Haiku with a mean difference of 7.8 ($p=0.052$), although this result is slightly above the conventional significance threshold, indicating a trend rather than a statistically significant difference. A qualitative review suggested that both human participants and LLMs most frequently erred in toxicology and multi-step trauma questions, whereas cardiology and infectious disease questions were associated with higher accuracy across groups.

Table 2. ANOVA and Levene's test results

Test	F	df1	df2	p
One-way ANOVA (Welch's)	31.3	9	20.8	<0.001
Homogeneity of variances test (Levene's)		9	60	0.017

ANOVA: Analysis of variance

Discussion

This study demonstrated that LLMs, particularly Claude Sonnet 3.5, outperformed most human groups who took the 50-question emergency medicine test administered by the TUS. The top-performing LLMs reached scores comparable to or higher than those of postgraduates with more than five years of experience.

Table 3. Post-hoc analysis (Games-Howell)

Group	MS5	MS6	PG1	PG5+	C4o	CS3.5	GA	C4o-m	GF	CH
MS5	-	-4.3 (0.48)	-9.3 (<0.001)	-14.4 (<0.001)	-15.5 (<0.001)	-17.3 (<0.001)	-14.5 (<0.001)	-12.5 (0.122)	-10.9 (0.004)	-7.7 (0.051)
MS6		-	-5.0 (0.273)	-10.1 (0.002)	-11.2 (<0.001)	-13.0 (<0.001)	-10.2 (0.002)	-8.2 (0.442)	-6.6 (0.141)	-3.4 (0.832)
PG1			-	-5.1 (0.062)	-6.2 (0.007)	-8.0 (<0.001)	-5.2 (0.04)	-3.2 (0.98)	-1.6 (0.993)	1.6 (0.994)
PG5+				-	-1.1 (0.985)	-2.9 (0.21)	-0.1 (1.0)	1.9 (0.999)	3.5 (0.595)	6.7 (0.091)
C-4o					-	-1.8 (0.211)	1.0 (0.968)	3.0 (0.98)	4.6 (0.256)	7.8 (0.052)
CS3.5						-	2.8 (0.168)	4.8 (0.817)	6.4 (0.094)	9.6 (0.028)
GA							-	2.0 (0.999)	3.6 (0.509)	6.8 (0.086)
C4o m								-	1.6 (1.0)	4.8 (0.884)
GF									-	3.2 (0.839)
CH										-

The values represent the mean difference (p value). MS5: Medical student-phase 5, MS6: Medical student-phase 6, PG1: Postgraduate one year, PG5+: Postgraduate 5+ years, C-4o: ChatGPT 4o, CS3.5: Claude Sonnet 3.5, GA: Gemini Advanced, C4o m: ChatGPT 4o Mini, GF: Gemini Flash, CH: Claude Haiku

Emergency medicine plays a pivotal role in managing acute, life-threatening conditions, demanding rapid decision-making and extensive clinical knowledge (11). Examinations such as the TUS ensure practitioners are prepared for high-pressure scenarios, highlighting the importance of valid training programs (1,12). The global shortage of emergency physicians highlights the need for innovative solutions to support training, especially in areas with limited access to experienced instructors (13). LLMs offer a potential solution to these gaps by providing accurate answers to complex questions. The performance of these methods in this study suggests that they could improve preparation for standardized tests, such as TUS. The frequent physiological changes during acute medical events highlight the need for accurate and easily accessible knowledge, which LLMs seem to provide effectively (14).

This study found that Claude Sonnet 3.5 outperformed most human groups, which is consistent with recent literature on AI in medical education. Roos et al. (15) reported that LLMs outperformed medical students on MCAT-style questions, suggesting their strength in knowledge-based assessments. A meta-analysis of 32 studies by Waldock et al. (16) confirmed that LLMs were consistently superior in general medical examinations, although specialty-specific research remains limited. A recent study from Japan by Akitomo et al. (17) revealed similar AI performance in dental board exams, supporting the current findings. However, Johri et al. (18) noted that LLMs sometimes struggle with context-specific reasoning in clinical scenarios; this challenge was not evident in this study, likely owing to the structured nature of the test questions. These results suggest that LLMs perform well in standardized settings but may face challenges in less structured environments (19,20).

Although a large number of pairwise comparisons were conducted, the primary aim of this study was not to interpret each contrast in isolation but to identify overarching performance patterns across participant groups and model types.

When the results are examined collectively, three consistent trends emerge. First, a marked performance gap is observed between undergraduate medical students and both LLMs and experienced emergency physicians. Second, performance convergence is evident between senior emergency physicians (PG5+) and top-performing LLMs, suggesting comparable performance on examination-based assessments. Third, substantial variability among LLMs themselves highlights the importance of model architecture and design choices in determining examination performance.

Given their consistency and high accuracy, advanced LLMs have the potential to be reliable tools for emergency medicine training. However, the variability among models, such as ChatGPT 4.0 Mini, suggests that not all LLMs are equally suitable for such applications, necessitating careful model selection for educational purposes (21). For human learners, these data highlight the importance of clinical experience for improving performance, as evidenced by the progression from MS5 to PG5+ (22).

The findings are likely applicable to academic settings with standardized emergency medicine examinations, particularly in urban tertiary institutions. The strong performance of LLMs suggests that they could enhance training by providing scalable educational resources, especially where access to instructors is limited (23,24). Their epidemiological and clinical importance lies in improving training efficiency, which could help increase the supply of qualified emergency medicine specialists.

The results indicate the potential of LLMs as educational tools in emergency medicine, based on their test performance. Their integration into medicine could support exam preparation and knowledge reinforcement. Further studies are needed to validate their clinical utility and address variability in responses.

Study Limitations

The study has several limitations that should be taken into account. The single-center design may limit its applicability to other educational settings. The controlled test environment does not replicate clinical pressures, potentially overestimating LLM performance. The variability in LLM responses across iterations indicates potential inconsistencies, which could affect reliability.

Limited qualitative analysis of response patterns restricts deeper insight into LLM reasoning processes. The absence of time pressure for LLMs may partially lead to an overestimation of their comparative performance. Given the large number of pairwise comparisons, the interpretation focused on contrasts with the greatest educational and clinical relevance rather than an exhaustive discussion of all individual differences.

Conclusion

This study demonstrates that LLMs outperform most human groups and perform comparably to experienced postgraduates on emergency medicine exam questions. These findings suggest that LLMs could be valuable educational tools for enhancing medical training in critical specialties. Integrating LLMs into educational programs may improve exam preparation and knowledge acquisition, particularly in resource-limited settings. Future research should investigate their clinical applications, address response variability, and validate findings across diverse contexts to ensure their effective integration into medical education.

Ethics

Ethics Committee Approval: This cross-sectional study was approved by the appropriate İstanbul Medipol University Non-Interventional Clinical Research Ethics Committee (decision number: 664, date: 04.07.2024).

Informed Consent: All participants provided informed consent prior to participation in the study.

Footnotes

Authorship Contributions: Surgical and Medical Practices - S.B.; Concept - Ş.Z.E.K.; Design - S.B.; Data Collection or Processing - S.B.; Analysis or Interpretation - S.B.; Literature Search - Ş.Z.E.K.; Writing - Ş.Z.E.K.

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. Koçak M, Oğuz AK, Akçalı Z. The role of artificial intelligence in medical education: an evaluation of large language models (LLMs) on the Turkish Medical Specialty Training Entrance Exam. BMC Med Educ. 2025; 25: 609.
2. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023; 2: e0000198.
3. Feigerlova E, Hani H, Hothersall-Davies E. A systematic review of the impact of artificial intelligence on educational outcomes in health professions education. BMC Med Educ. 2025; 25: 129.
4. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med. 2023; 6: 120.
5. Preiksaitis C, Ashenburg N, Bunney G, Chu A, Kabeer R, Riley F, et al. The role of large language models in transforming emergency medicine: scoping review. JMIR Med Inform. 2024; 12: e53787.
6. Knoedler L, Alfertshofer M, Knoedler S, Hoch CC, Funk PF, Cotofana S, et al. Pure wisdom or potemkin villages? A comparison of ChatGPT 3.5 and ChatGPT 4 on USMLE step 3 style questions: quantitative analysis. JMIR Med Educ. 2024; 10: e51148.
7. Morjaria L, Burns L, Bracken K, Ngo QN, Lee M, Levinson AJ, et al. Examining the threat of ChatGPT to the validity of short answer assessments in an undergraduate medical program. J Med Educ Curric Dev. 2023; 10: 23821205231204178.
8. Meral G, Ateş S, Günay S, Öztürk A, Kuşdoğan M. Comparative analysis of ChatGPT, Gemini and emergency medicine specialist in ESI triage assessment. Am J Emerg Med. 2024; 81: 146-50.
9. Lucas HC, Upperman JS, Robinson JR. A systematic review of large language models and their implications in medical education. Med Educ. 2024; 58: 1276-85.
10. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature. 2023; 620: 172-80.
11. Vella KM, Hall AK, van Merrienboer JJG, Hopman WM, Szulewski A. An exploratory investigation of the measurement of cognitive load on shift: application of cognitive load theory in emergency medicine. AEM Educ Train. 2021; 5: e10634.
12. Lombardi CV, Chidiac NT, Record BC, Laukka JJ. USMLE Step 1 and Step 2 CK as indicators of resident performance. BMC Med Educ. 2023; 23: 543.
13. Ayoola AS, Acker PC, Kalanzi J, Strehlow MC, Becker JU, Newberry JA. A qualitative study of an undergraduate online emergency medicine education program at a teaching Hospital in Kampala, Uganda. BMC Med Educ. 2022; 22: 84.
14. Williams CYK, Zack T, Miao BY, Sushil M, Wang M, Kornblith AE, et al. Use of a large language model to assess clinical acuity of adults in the emergency department. JAMA Netw Open. 2024; 7: e248895.
15. Roos J, Kasapovic A, Jansen T, Kaczmarczyk R. Artificial intelligence in medical education: comparative analysis of ChatGPT, Bing, and medical students in Germany. JMIR Med Educ. 2023; 9: e46482.
16. Waldock WJ, Zhang J, Guni A, Nabeel A, Darzi A, Ashrafian H. The accuracy and capability of artificial intelligence solutions in health care examinations and certificates: systematic review and meta-analysis. J Med Internet Res. 2024; 26: e56532.
17. Akitomo T, Hamada M, Tsuge Y, Kaneki A, Ogawa M, Nishimura T, et al. Artificial intelligence's performance on the Japanese National Dental Examination. Cureus. 2024; 16: e73103.
18. Johri S, Jeong J, Tran BA, Schlessinger DI, Wongvibulsin S, Barnes LA, et al. An evaluation framework for clinical use of large language models in patient interaction tasks. Nat Med. 2025; 31: 77-86.
19. Sonoda Y, Kurokawa R, Hagiwara A, Asari Y, Fukushima T, Kanzawa J, et al. Structured clinical reasoning prompt enhances LLM's diagnostic capabilities in diagnosis please quiz cases. Jpn J Radiol. 2025; 43: 586-92.
20. Watanabe T, Baba A, Fukuda T, Watanabe K, Woo J, Ojiri H. Role of visual information in multimodal large language model performance: an evaluation using the Japanese nuclear medicine board examination. Ann Nucl Med. 2025; 39: 217-24.
21. Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman H, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. JAMA Netw Open. 2024; 7: e2440969.
22. Davis D, Galbraith R; American College of Chest Physicians Health and Science Policy Committee. Continuing medical education effect on practice performance: effectiveness of continuing medical education: American College of Chest Physicians Evidence-Based Educational Guidelines. Chest. 2009; 135: 42S-8S.
23. Atsukawa N, Tatekawa H, Oura T, Matsushita S, Horiuchi D, Takita H, et al. Evaluation of radiology residents' reporting skills using large language models: an observational study. Jpn J Radiol. 2025; 43: 1204-12.
24. Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, et al. Large language models in medical education: opportunities, challenges, and future directions. JMIR Med Educ. 2023; 9:e48291.

İstanbul Med J 2026; 27(2): 103-11
DOI: 10.4274/imj.galenos.2026.00018

Evaluation of Balance and Core Endurance in Patients with Fibromyalgia Syndrome

Songul Baglan Yentur[1], Muhammet Sahin Elbasti[2], Bekir Dagdeviren[1], Derya Cetintas[2]

[1]Fırat University Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, Elazığ, Türkiye
[2]Elazığ Fethi Sekin City Hospital, Clinic of Physical Medicine and Rehabilitation, Elazığ, Türkiye

ABSTRACT

Introduction: This study was designed to examine disparities in core endurance and balance among patients with fibromyalgia syndrome (FMS) relative to healthy controls.

Methods: A case-control design was used, including 44 FMS patients and 44 matched controls. Assessments included the McGill tests for core endurance; the HUR BTG4 Balance Master System and the Berg Balance Scale (BBS) for balance; and the Fibromyalgia Impact Questionnaire for functional status.

Results: There were no significant demographic differences between the groups ($p>0.05$). However, FMS patients exhibited significantly lower scores on the BBS ($p=0.017$), trunk extension ($p<0.001$), and prone bridge tests ($p<0.001$) than controls. Significant impairments were observed in all static and dynamic balance parameters, including SP-COG-EO area ($p<0.001$) and USP-COG-EC velocity ($p<0.001$). A moderate negative correlation was found between the right-side bridge test and velocity during the eyes-closed balance test (r: -0.309, $p=0.042$).

Conclusion: Core endurance and balance are significantly impaired in patients with FMS. While individual deficits exist, the lack of strong correlation between most endurance and balance measures suggests these impairments may stem from independent pathophysiological mechanisms in FMS.

Keywords: Balance, core endurance, fibromyalgia syndrome

Introduction

Fibromyalgia syndrome (FMS) is a long-term condition of unknown origin, primarily affecting women aged 40 to 60. The exact causes and underlying mechanisms of FMS remain uncertain (1). Pain, stiffness, subjective soft-tissue swelling, fatigue, sleep disturbance, and psychiatric and cognitive disorders are among the symptoms observed in patients with FMS (2). However, balance problems are among the most frequent and disturbing symptoms in patients with FMS (3). A study of 2596 FMS patients reported that balance impairment was one of the 10 most distressing symptoms in patients' self-reported statements (4). A significant proportion of patients (45%) experience balance problems which may be attributable to multiple factors, including sensory and cognitive deficits and symptoms such as pain, poor sleep, fatigue, and muscle weakness (5,6). In a study investigating balance in FMS, 68% of 486 patients reported balance problems (7).

Balance is crucial for performing everyday tasks, such as walking, running, and climbing stairs (8). Postural problems affecting the vertebral column seen in FMS affect the spatiotemporal parameters of walking and increase the risk of falls (9). The core is a musculoskeletal structure conceptualized as a cylinder (10). Studies demonstrated that core stabilization training improved static and dynamic balance (11,12). In addition, balance training should also include core stabilization exercises as it improves body awareness and acts as a muscular corset that stabilizes the trunk with or without limb movement (13). Generally, the purpose of the core muscles is to stabilize the spine during functional movements while the body seeks to maximize balance. It has been claimed that improving core stability will increase dynamic stability. The body's ability to maintain dynamic stability requires neuromuscular control of all segments, including both proximal and distal joints. The body's core responds to disturbances in dynamic stability by utilizing the core musculature and transferring forces through the trunk (14). To the best of our knowledge, research specifically assessing core stability in individuals with FMS is extremely limited. The primary contribution of this study is its comprehensive approach: it is the first to evaluate both static and dynamic core endurance alongside objective balance metrics measured using the HUR BTG4 Balance Master System. By using an

***This paper was presented as an abstract at EULAR 2025, Barcelona, Spain.**

Address for Correspondence: Assoc. Prof., Songül Bağlan Yentür, MD, Fırat University Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, Elazığ, Türkiye
E-mail: songulbaglan23@hotmail.com **ORCID ID:** orcid.org/0000-0001-9394-4817

Cite this article as: Baglan Yentur S, Elbasti MS, Dagdeviren B, Cetintas D. Evaluation of balance and core endurance in patients with fibromyalgia syndrome. İstanbul Med J. 2026; 27(2): 103-11

Received: 15.08.2025
Accepted: 13.02.2026
Publication Date: 12.05.2026

objective platform to measure sway area, length, and velocity, this study fills a critical gap in the literature by examining the direct relationship between trunk stabilization and balance impairments in FMS. This objective data is vital for determining whether core endurance is a primary driver of balance dysfunction or other sensory-cognitive factors play a dominant role, thereby guiding more targeted rehabilitation interventions.

While previous research has explored the connection between balance issues and core stability, there appears to be a lack of studies specifically assessing core stability in individuals with FMS, to the best of the authors' knowledge. The main objective of this study was to examine and compare core stability and balance in FMS patients and healthy individuals. A secondary goal was to explore the potential relationship between core stability and balance impairments in individuals with FMS.

Methods

This study enrolled 44 patients diagnosed with FMS from the outpatient clinic of Elazığ Fethi Sekin City Hospital and 44 age- and sex-matched healthy volunteers. The sample size was determined a priori using G*Power software (version 3.1.9.4). Based on the effect size (d: 0.60) derived from the mean difference in eyes closed-center of gravity (EC-COG) velocity scores reported by Tas et al. (15), a power analysis with α: 0.05 and power (1-β): 0.80 indicated that were required to detect an effect size of d: 0.60. The study protocol received approval from the Fırat University Non-Interventional Clinical Research Ethics Committee (decision number: 2022/11-24, date: 06.10.2022), and written informed consent was obtained from all participants prior to their inclusion. Inclusion criteria encompassed individuals who were aged between 18 and 65, who met the clinical diagnostic criteria for FMS, who had been on a stable medication regimen for at least three months, and who were able to follow complex motor instructions required for the assessments. Exclusion criteria included severe spinal stenosis, concurrent systemic inflammatory rheumatic conditions, pregnancy, cancer, active infections, surgical procedures within the past three months, or any physical limitations that could interfere with the administration of core endurance assessments.

Outcome Measurements

The demographic details of participants—including age, height, weight, occupation, medication use, and smoking status—were collected at the outset. To assess core endurance, both static and dynamic tests were administered. Static core endurance was evaluated using the side bridge, trunk flexion, trunk extension, and prone bridge tests. To assess dynamic core endurance, participants performed modified push-up and sit-up tests. Functional capacity in individuals with fibromyalgia was measured using the Fibromyalgia Impact Questionnaire (FIQ). Balance, both static and dynamic, was assessed using the HUR BTG4 Balance Master System and the Berg Balance Scale (BBS).

McGill Core Endurance Tests

Prior to actual testing, participants performed a single familiarization trial for each McGill test to ensure correct technique. Each assessment were performed once to prevent fatigue. The test completion criterion (end-point) was defined as the moment when the participant could no longer maintain the required isometric position or when they deviated from the neutral spinal alignment despite verbal cues.

Side-bridge test: Participants were asked to lift their hips off the floor in a side-lying position, supported on the elbow and forearm. After the patients positioned their whole bodies in a straight line, supported on their feet and elbows, timing began with a stopwatch. The time maintained for the right and left sides was recorded as a score (16).

Trunk flexion test: Participants were asked to position the upper trunk at 60º flexion, with the hips and knees at 90º flexion. The duration for which patients could maintain this angle while standing was recorded (16).

Trunk extension test: Participants were asked to lie prone with the upper part of the iliac crest positioned beyond the edge of the bed, with their hips and knees fully extended. The time that the horizontal position with their arms in the inverted-T position was maintained was measured with a stopwatch and recorded (16).

Prone bridge test: Participants were asked to support themselves on their elbows and toes while lying prone and to maintain a horizontal position on the floor. The duration for which the position was maintained was recorded (16).

Modified push-ups test: Participants were asked to assume the prone position, cross their feet, and place their hands on an exercise mat. They were then asked to lift their trunk by extending their elbows, without disturbing the straight alignment of the trunk and hips. The number of repetitions performed within 30 seconds was recorded as a score (16).

Sit-ups test: Participants were asked to place their hands at the sides of their heads and perform trunk flexion while in the supine position with their knees at 90° of flexion. The number of repetitions performed within 30 seconds was recorded as a score (16).

Balance Assessments

Balance was assessed using the HUR BTG4 Balance Master System® (HUR International, Finland) on both stable and unstable platforms, as well as the BBS. The system utilizes HUR Smart Balance Software, which is specifically developed for balance evaluation and training purposes. Before each assessment, the stability of both platform types was confirmed. Data were recorded at a sampling rate of 100 Hz and then filtered. The primary focus of this study was the 95% confidence ellipse area of the center of pressure for analyzing balance. This equipment is particularly useful for evaluating postural control and developing individualized rehabilitation strategies (Figure 1) (17).

Before starting the assessment, participants stood still for at least five seconds. They then stood upright with their arms relaxed at their sides. Measurements were obtained under two visual conditions—eyes open (EO) and EC, the latter eliminating visual input, and on a foam surface to reduce proprioceptive feedback, allowing the evaluation of static balance. Body sway was analyzed to assess the contributions of the visual, vestibular, and proprioceptive sensory systems (see Figures 2-4). During testing, participants stood barefoot on the platform with their heels positioned 2 cm apart and

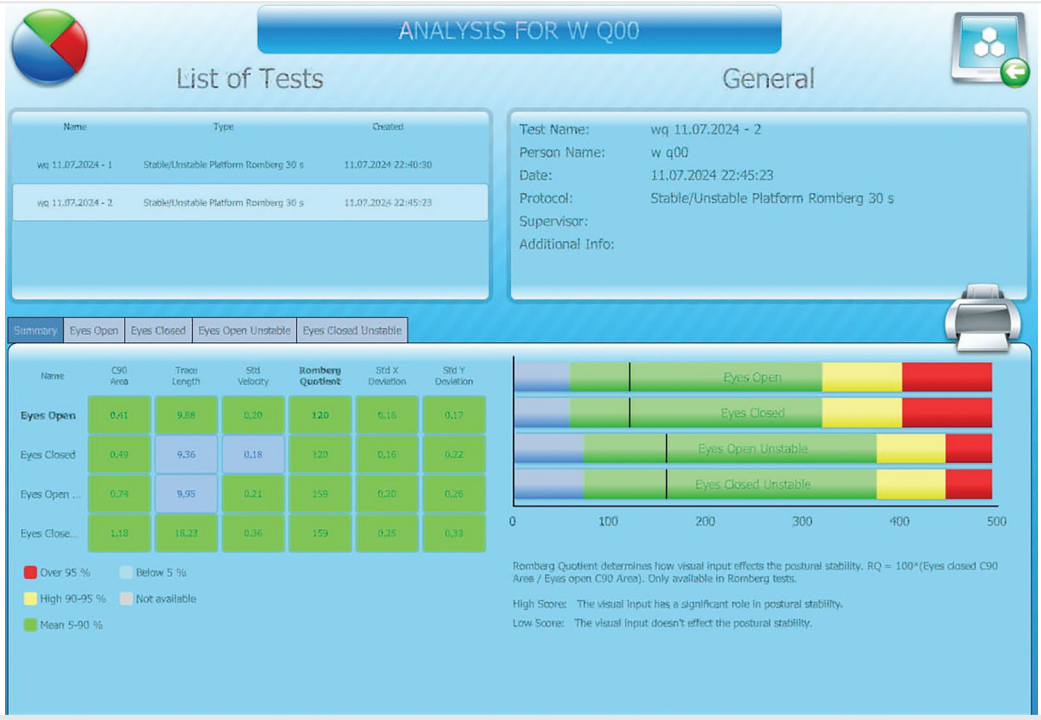

Figure 1. Balance assessments on stable an unstable platforms

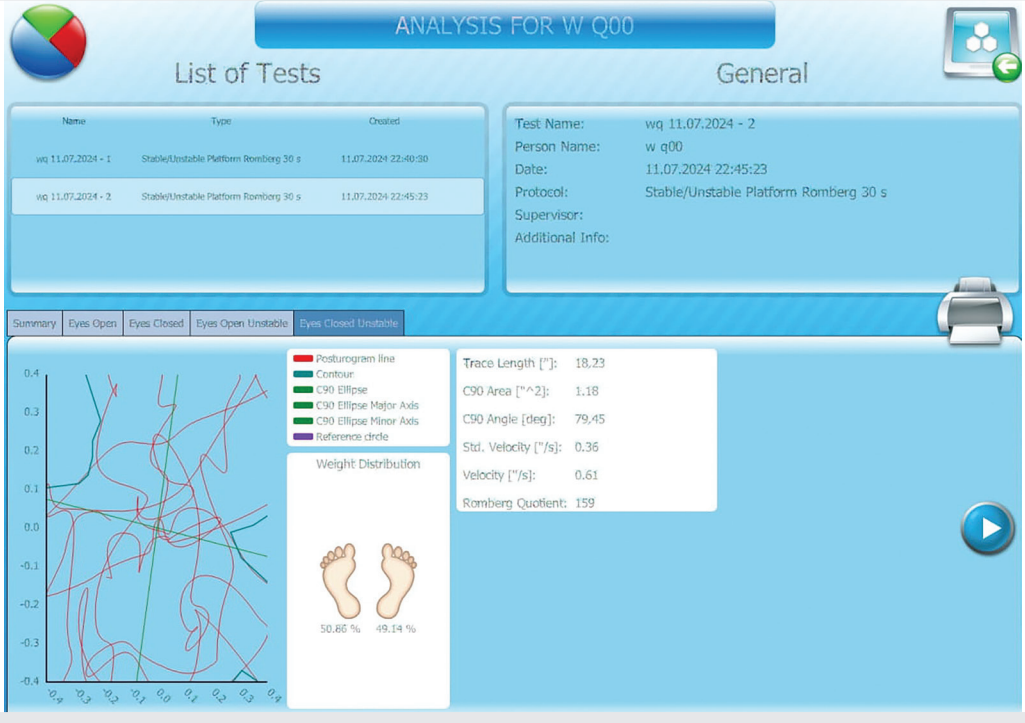

Figure 2. Eyes closed balance evaluation on stable platform

arms at their sides for 30 seconds. Each test was followed by a 60-second rest period. This protocol examines proprioceptive function, recognizing that when proprioception is impaired, balance relies more heavily on the visual and vestibular systems. However, in cases of proprioceptive dysfunction, closing the eyes disrupts this compensatory process, resulting in balance difficulties.

To assess dynamic balance, participants experienced controlled perturbations from the right, left, and horizontal directions, with peak accelerations reaching 9.6 m/s². Each individual's leaning angle was determined by considering their height and the shift in their COG from a neutral stance. The evaluation involved asking participants to lean forward, backward, to the right, and to the left to assess their balance responses.

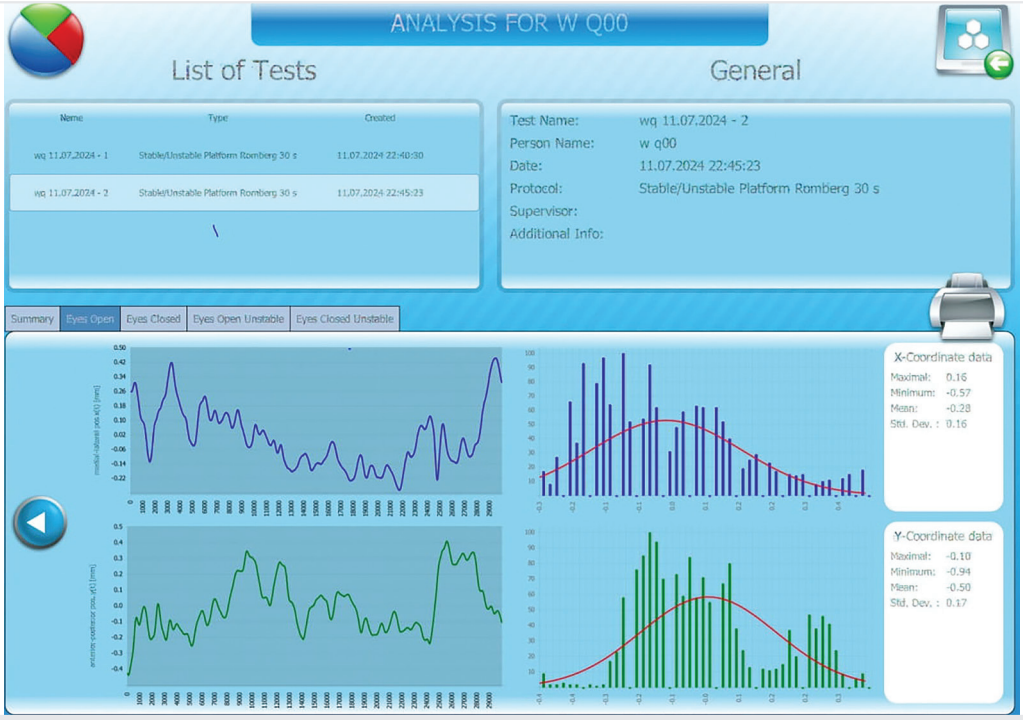

Figure 3. Balance assessments with eyes open positions on stable platform

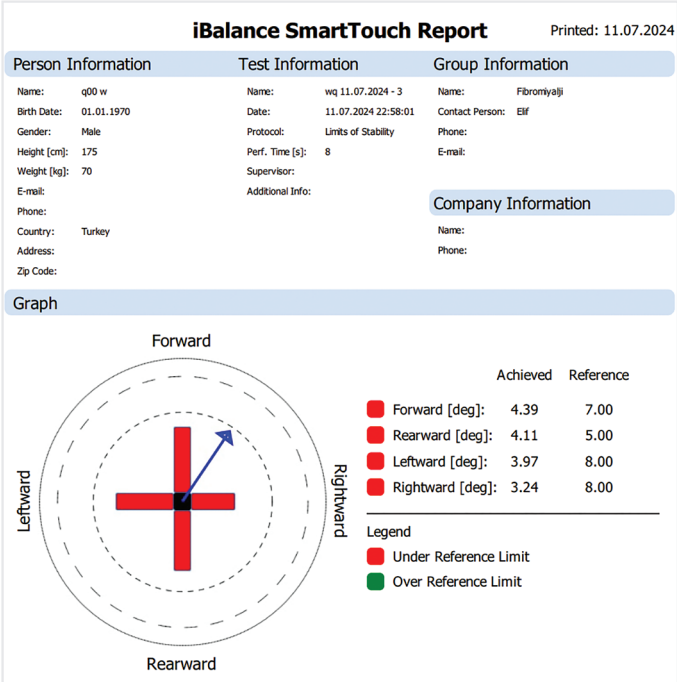

Figure 4. Dynamic balance assessment

Berg Balance Scale

Balance was assessed using the BBS. The BBS is a reliable and precise tool designed to evaluate balance. It comprises 14 tasks that assess an individual's capacity to maintain stability during various movements and postural shifts. Each task is designed to assess how well an individual can complete it independently and within a specified time or over a specified distance. Scores for each item range from 0 to 4 points (18).

Functional Evaluation

Patient function was evaluated using the FIQ. The total FIQ score is derived from the sum of its ten items. Item one assesses daily activities using a Likert scale anchored at 0 ("always able to do") and 3 ("never able to do") (19).

Statistical Analysis

Statistical analyses were conducted using the Statistical Package for the Social Sciences (SPSS) for Windows, version 22.0. Descriptive statistics were presented as frequencies, percentages, medians, and interquartile ranges. The normality of continuous data was assessed using the Kolmogorov-Smirnov and Shapiro-Wilk tests. For data that were not normally distributed, the nonparametric Mann-Whitney U test was applied. Categorical variables were compared between groups using the Pearson chi-square test. Correlations were examined using the Spearman correlation coefficient. Statistical significance was set at $p<0.05$.

Results

This study compared 44 patients with FMS to 44 age- and sex-matched healthy controls. The groups were statistically similar in their sociodemographic profiles (Table 1).

Significant differences were observed in several measures. The FMS group demonstrated markedly lower performance on the BBS ($p=0.017$) and specific static core endurance tests, including trunk extension ($p<0.001$) and the prone bridge ($p<0.001$) (Table 2). Furthermore, both static and dynamic balance metrics (e.g., SP-COG-EC area, $p<0.001$; USP-COG-EC velocity, $p<0.001$) were significantly impaired in the patient group compared to controls (Table 3). However, not all tests showed

Table 1. Demographics and characteristics of FMS patients and healthy controls				
Parameters [median (IQR: 25/75)]		**FMS group (n=44)**	**Healthy controls (n=44)**	**p**
Age/year		45.00 [43.25; 51.00]	45.00 [39.25; 51.00]	0.538
Weight/kg		70.00 [63.00; 74.75]	72,00 [65.00; 77.00]	0.448
Height/cm		160.00 [160.00; 163.00]	160.50 [154.25; 165.00]	0.285
BMI/kg/cm^2		27.34 [24.40; 27.82]	26.99 [25.42; 31.20]	0.324
Disease duration/year		2.00 [2.00; 4.00]		
FIQ		66.44 [62.05; 74.86]		
Medication (%)	Myorelaxan	6 (50%)	1 (25%)	0.313
	HT	3 (25%)	1 (25%)	
	Diabetes	2 (16.7%)	0 (0%)	
	Antidepressant	1 (8.3%)	1 (25%)	
	Goiter	0 (0%)	1 (25%)	
Smoking (%)	Yes	14 (31.8%)	10 (22.7%)	0.338
	No	30 (68.2%)	34 (77.3%)	
Occupation (%)	Yes	12 (27.3%)	13 (29.5%)	0.813
	No	32 (72.7%)	31 (70.5%)	

IQR: Interquartile range, FMS: Fibromyalgia syndrome, BMI: Body mass index, FIQ: Fibromyalgia Impact Questionnaire, HT: Hypertension

Table 2. Comparison of BBS and core endurance tests between FMS group and healthy controls			
Parameters [median (IQR: 25/75)]	**FMS group (n=44)**	**Healthy controls (n=44)**	**p**
BBS	48.00 [48.00; 49.75]	51.50 [46.25; 55.00]	**0.017**
Side bridge test-R	6.64 [3.25; 9.75]	8.00 [4.00; 10.00]	0.366
Side bridge test-L	6.09 [3.00; 8.50]	6.85 [3.25; 10.00]	0.402
Trunk flexion test	11.37 [3.00; 17.75]	14.00 [10.25; 16.00]	0.137
Trunk extension test	8.38 [5.25; 9.75]	16.00 [11.50; 19.75]	**<0.001**
Prone bridge test	6.37 [4.00; 7.00]	13.00 [9.25; 18.00]	**<0.001**
Push-ups test	6.00 [5.00; 8.46]	7.00 [5.00; 8.00]	0.956
Sit-ups test	7.00 [3.50; 8.85]	8.00 [6.00; 9.00]	0.339

IQR: Interquartile range, FMS: Fibromyalgia syndrome, BBS: Berg Balance Scale, R: Right, L: Left

differences; performance on dynamic core endurance tests and certain static tests (side bridge and trunk flexion) was comparable between the groups (Table 2).

Within the FMS group, correlation analyses revealed specific relationships. A moderate inverse correlation was found between performance on the right-side bridge test and velocity during the eyes-closed balance test on a stable platform(r: -0.309, p=0.042) (Table 4). Positive correlations were identified between the BBS and two parameters of EC balance on an unstable platform: sway area (r: 0.328, p=0.030)and sway length (r: 0.329, p=0.029) (Table 5).

Discussion

This study aimed to compare core endurance and balance between individuals with FMS and a healthy control group. A secondary objective was to examine the relationship between core endurance and balance in those with FMS. The findings revealed that self-perceived balance differed significantly between FMS patients and healthy controls matched for age and gender. Additionally, the trunk extension and prone bridge assessments showed notable differences in static core endurance. Significant group differences were also identified in balance tests conducted with EO and with EC on both stable and unstable surfaces using the HUR BTG4 Balance Master System® (HUR International, Finland). Analysis of the association between core endurance and balance parameters revealed a significant correlation between the eyes-closed velocity test and the right-side bridge test, performed on a stable surface. When the relationship between balance tests performed on an unstable surface and core endurance tests was evaluated, a relationship was observed between the self-reported balance measurement and the EC area and EC length measurements. It was concluded that FMS patients had impaired balance and performed worse on some static core endurance tests than an age- and gender-matched healthy control group.

Many studies have concluded that FMS patients have impaired balance. It was observed at a frequency between 45% and 68% and reported to be among the ten most challenging symptoms. Balance impairment in FMS patients may also increase the risk of falls (20). The incidence of falls has been reported as 1.75 per 6-month period (9). Falls are associated with postural instability, balance, and executive function and processing speed (21). Collado-Mateo et al. (9) found that FMS may cause fear of falling, balance problems, an increased number of falls.

Table 3. Comparison of static and dynamic balance tests between FMS group and healthy controls

Parameters [Median (IQR: 25/75)]	FMS Group (n=44)	Healthy controls (n=44)	p
SP-COG-EO-area	0.41 [0.27; 0.41]	0.19 [0.12; 0.26]	**<0.001**
SP-COG-EO-length	13.34 [7.64; 13.53]	7.05 [5.85; 7.76]	**<0.001**
SP-COG-EO-velocity	0.19 [0.14; 0.19]	0.15 [0.13; 0.16]	**0.004**
SP-COG-EC-area	0.96 [0.37; 0.96]	0.23 [0.12; 0.35]	**<0.001**
SP-COG-EC-length	13.53 [13.07; 14.71]	9.29 [7.24; 11.08]	**<0.001**
SP-COG-EC-velocity	0.30 [0.25; 0.30]	0.20 [0.17; 0.26]	**<0.001**
USP-COG-EO-area	0.66 [0.48; 0.66]	0.42 [0.24; 0.62]	**0.003**
USP-COG-EO-length	11.48 [10.19; 11.48]	10.13 [7.90; 10.80]	**0.003**
USP-COG-EO-velocity	0.27 [0.24; 0.27]	0.23 [0.17; 0.25]	**0.007**
USP-COG-EC-area	1.31 [0.77; 1.31]	0.67 [0.33; 0.85]	**<0.001**
USP-COG-EC-length	22.35 [17.48; 22.35]	14.85 [10.66; 16.35]	**<0.001**
USP-COG-EC-velocity	0.82 [0.42; 0.82]	0.30 [0.22; 0.38]	**<0.001**
Forward	2.63 [1.80; 3.80]	3.64 [2.69; 4.49]	**0.009**
Backward	2.86 [2.53; 3.10]	4.48 [3.64; 5.02]	**<0.001**
Left	2.99 [2.89; 3.34]	4.25 [3.45; 4.89]	**<0.001**
Right	2.84 [2.45; 3.45]	4.85 [3.86; 5.84]	**<0.001**

FMS: Fibromyalgia syndrome, IQR: Interquartile range, SP: Stable platform, COG: Center of gravity, EO: Eyes open, EC: Eyes closed, USP: Unstable platform

Table 4. Correlations between static balance measurements and core endurance tests and BBS in FMS group

Parameters		BBS	FIQ	Side bridge test-R	Side bridge test-L	Trunk flexion test	Trunk extansion test	Prone bridge test	Push-ups test	Sit-ups test
BBS	r	1	-0.263	0.255	0.112	-0.051	0.148	0.231	-0.182	-0.188
	p		0.085	0.095	0.470	0.743	0.339	0.131	0.237	0.223
SP-COG-EO-area	r	0.060	-0.002	-0.184	-0.166	-0.094	0.126	-0.121	-0.124	-0.003
	p	0.700	0.992	0.232	0.281	0.544	0.416	0.433	0.424	0.986
SP-COG-EO-length	r	0.258	-0.025	-0.072	0.034	-0.168	0.284	0.076	0.023	-0.047
	p	0.090	0.874	0.641	0.826	0.276	0.062	0.622	0.881	0.761
SP-COG-EO-velocity	r	0.221	-0.063	-0.218	-0.151	-0.237	-0.070	-0.112	0.010	-0.130
	p	0.149	0.686	0.156	0.328	0.122	0.652	0.468	0.947	0.401
SP-COG-EC-area	r	0.018	-0.040	-0.058	0.016	-0.088	-0.011	-0.007	0.020	0.220
	p	0.908	0.796	0.709	0.918	0.570	0.942	0.963	0.193	0.151
SP-COG-EC-length	r	0.034	0.082	-0.276	-0.203	-0.225	-0.230	-0.279	0.028	-0.006
	p	0.827	0.595	0.070	0.185	0.142	0.133	0.067	0.856	0.970
SP-COG-EC-velocity	r	0.039	-0.014	-0.309	-0.225	-0.236	-0.119	-0.216	0.005	-0.020
	p	0.804	0.926	**0.042**	0.141	0.124	0.441	0.158	0.974	0.897

FMS: Fibromyalgia syndrome, BBS: Berg Balance Scale, COG: Center of gravity, SP: Stable platform, EO: Eyes open, EC: Eyes closed, FIQ: Fibromyalgia Impact Questionnaire, R: Right, L: Left

It has also been argued that balance problems may be associated with fear of falling (9). Cognitive functions are known to be affected in patients with FMS. Increased cognitive problems may elevate the risk of balance impairment and falls (22). In our study, we evaluated the balance of patients with FMS on both stable and unstable surfaces and found significant differences compared with healthy controls. Evaluating balance using the HUR BTG4 Balance Master System allowed us to obtain more objective results. Visual, vestibular, and proprioceptive sensations were also evaluated using this test. Area, length, and velocity measurements were evaluated on stable and unstable platforms with EO and EC. Across all evaluated parameters, patients had higher values than healthy subjects, suggesting greater oscillation to maintain balance and, therefore, poorer postural control. In addition, we found a significant difference on the BBS, which is a performance-based measure of balance administered by an examiner. Consistent with the literature, these results demonstrate that patients with FMS may have impaired balance. Tas et al. (15) evaluated balance in patients with FMS and in healthy controls using the same device; however, unlike our study, they did not find significant differences across many balance parameters. The same study found a significant difference in BBS scores between patients with FMS and healthy controls, which is similar to our findings.

Table 5. Correlations between dynamic balance measurements and core endurance tests and BBS in FMS group

Parameters		BBS	FIQ	Side bridge test-R	Side bridge test-L	Trunk flexion test	Trunk extansion test	Prone bridge test	Push-ups test	Sit-ups test
USP-COG-EO-area	r	-0.224	0.151	-0.195	-0.121	-0.016	0.175	-0.041	0.051	0.183
	p	0.143	0.328	0.205	0.435	0.918	0.255	0.791	0.743	0.235
USP-COG-EO-length	r	-0.075	0.017	-0.002	-0.003	0.115	0.185	0.034	0.079	0.064
	p	0.630	0.913	0.991	0.982	0.459	0.229	0.827	0.610	0.682
USP-COG-EO-velocity	r	0.003	-0.137	0.034	0.019	0.085	0.250	0.023	0.033	-0.060
	p	0.983	0.376	0.827	0.902	0.583	0.102	0.883	0.833	0.698
USP-COG-EC-area	r	0.328*	-0.244	-0.005	0.059	-0.064	0.070	0.055	0.067	0.019
	p	**0.030**	0.110	0.974	0.703	0.680	0.652	0.722	0.665	0.905
USP-COG-EC-length	r	0.329*	-0.272	0.018	0.074	-0.074	0.057	0.013	0.113	-0.091
	p	**0.029**	0.074	0.906	0.634	0.635	0.713	0.935	0.466	0.555
USP-COG-EC-velocity	r	0.091	-0.144	-0.032	0.091	-0.133	0.186	0.110	0.104	-0.032
	p	0.559	0.350	0.837	0.555	0.388	0.227	0.478	0.502	0.834

FMS: Fibromyalgia syndrome, BBS: Berg Balance Scale, COG: Center of gravity, USP: Unstable platform, EO: Eyes open, EC: Eyes closed, R: Right, L: Left

In this study, dynamic and static core endurance tests were performed on patients with FMS. The results of the study showed a significant difference between FMS patients and healthy controls in the prone bridge test and the trunk extension test. Although no significant difference was obtained, dynamic and static core endurance test scores were lower in patients. This indicated that core endurance was lower in patients with FMS than in healthy controls. In this study, symptoms such as pain, exercise intolerance, and decreased functional capacity in patients with FMS may cause a decrease in core endurance. Increased pain during activity can lead to pain-induced reflex inhibition of muscles, limiting the duration for which a patient can sustain a muscle contraction. Our study assessed muscle endurance by requiring participants to sustain a muscle contraction for an extended period (20). Therefore, the decreased performance observed on extensor and side-bridge endurance tests among FMS patients may be related to impaired muscle microcirculation during both static and dynamic activities. Previous research has suggested that reduced aerobic endurance in individuals with fibromyalgia may result from reduced blood flow and impaired oxygen delivery to muscle tissue. These changes may be associated with atrophy of type-II muscle fibers and increased mitochondrial content in type-I fibers, leading to disrupted oxidative metabolism and reduced ATP synthesis, both of which manifest as muscle weakness and fatigue (23). Additionally, in FM muscles, the combination of decreased phosphorylation potential, low oxidative capacity, and an inefficient work-to-energy-cost ratio may lead to a noticeable decline in endurance (23). This may be the mechanism by which core endurance was lower in patients with FMS than in healthy controls in our study. Studies that evaluate all symptoms are needed to determine the cause of decreased core endurance. This study focused on core endurance and balance in patients with FMS and healthy controls. Significant differences in trunk extension and prone-bridge tests were found in our study. High levels of fatigue are observed in the multifidus muscle during the trunk extension test. This is because the multifidus muscle is overactive. The flexed posture seen in patients with FMS may have led to a greater weakening of the back muscles. More detailed investigations can be conducted to yield clearer results. To the authors' knowledge, only one study has evaluated core endurance in FMS patients: Sindwani and Kaur (24). However, balance and its relationship with core endurance were not evaluated in their study. In addition, static core endurance tests were evaluated in their study, whereas dynamic core endurance tests were not. However, similar to our study, it was concluded that core endurance test scores were lower than those of healthy controls.

The relationship between dynamic and static core endurance tests and balance tests on stable and unstable platforms was investigated in this study. On the stable surface, the eyes-closed velocity test and the right side-bridge test showed a moderately significant negative relationship, but no significant relationships were found among the other parameters. This result is surprising. It was concluded that core endurance did not seem to be among the causes of impaired balance in FMS patients, whose core endurance tests and balance were found to be worse than those of the healthy control group. More patients and more detailed studies are needed. Similarly, no significant relationship was found between core endurance and balance in female athletes (25). Gordon et al. (26) also did not find a significant relationship between balance and core endurance in their study of lacrosse players. In both studies, balance was evaluated with the Star Excursion Balance Test. However, a weak relationship was found between core endurance and balance in dancers. The reason was attributed to a possibly greater proportion of type II fibers in dancers (27). More research is needed to clarify the relationship between core endurance and balance in patients with FMS.

The functional status of patients with FMS was evaluated using the FIQ in this study. Based on the evaluation results, no significant relationship was observed between FIQ and the balance and core endurance tests. Although this study concluded that patients who participated in our

study had balance impairment, it was predicted that the impairment was not at a level that would affect independence in daily life.

Study Limitations

Our study had some limitations. That we did not evaluate pain and activities of daily living is a limitation. However, functional assessment was performed, including pain, blood parameters, vitamin levels, and activities of daily living. Nevertheless, different results can be obtained with a more detailed evaluation. Secondly, our study included only women and did not include men, which is a limitation. Third, we did not evaluate the fall frequency or fear of falling among the patients who participated in the study. If future studies are designed to include these parameters, more comprehensive results can be obtained.

Conclusion

In conclusion, this study aimed to compare core endurance and balance between individuals with FMS and healthy controls, as well as to explore the potential relationship between these two factors. The findings indicated that patients with FMS exhibited both balance impairments and reduced core endurance relative to the control group, consistent with existing literature. However, no significant correlation was observed between core endurance and balance. This research is noteworthy as it represents the first attempt to specifically investigate the connection between core endurance and balance in FMS patients. Further research is recommended to explore this relationship in greater depth and to emphasize the role of balance and core stability exercises in managing FMS.

Ethics

Ethics Committee Approval: The study protocol received approval from the Fırat University Non-Interventional Clinical Research Ethics Committee (decision number: 2022/11-24, date: 06.10.2022).

Informed Consent: The written informed consent was obtained from all participants prior to their inclusion.

Footnotes

Authorship Contributions: Concept - S.B.Y.; Design - S.B.Y.; Data Collection or Processing - B.D., D.Ç.; Analysis or Interpretation - S.B.Y., M.Ş.E., B.D., D.Ç.; Literature Search - S.B.Y., M.Ş.E., B.D., D.Ç.; Writing - S.B.Y., M.Ş.E.

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. Bellato E, Marini E, Castoldi F, Barbasetti N, Mattei L, Bonasia DE, et al. Fibromyalgia syndrome: etiology, pathogenesis, diagnosis, and treatment. Pain Res Treat. 2012; 2012: 426130.
2. Cassisi G, Sarzi-Puttini P, Alciati A, Casale R, Bazzichi L, Carignola R, et al. Symptoms and signs in fibromyalgia syndrome. Reumatismo. 2008; 60(Suppl 1): 15-24.
3. Pérez-de-Heredia-Torres M, Huertas-Hoyas E, Martínez-Piédrola R, Palacios-Ceña D, Alegre-Ayala J, Santamaría-Vázquez M, et al. Balance deficiencies in women with fibromyalgia assessed using computerised dynamic posturography: a cross-sectional study in Spain. BMJ Open. 2017; 7: e016239.
4. Bennett RM, Jones J, Turk DC, Russell IJ, Matallana L. An internet survey of 2,596 people with fibromyalgia. BMC Musculoskelet Disord. 2007; 8: 27.
5. Muto LH, Sauer JF, Yuan SL, Sousa A, Mango PC, Marques AP. Postural control and balance self-efficacy in women with fibromyalgia: are there differences? Eur J Phys Rehabil Med. 2014; 51: 149-54.
6. Trevisan DC, Driusso P, Avila MA, Gramani-Say K, Moreira FMA, Parizotto NA. Static postural sway of women with and without fibromyalgia syndrome: a cross-sectional study. Clin Biomech (Bristol). 2017; 44: 83-9.
7. Katz R, Ferbert S, Leavitt F. Fibromyalgia patients report many symptoms other than pain and fatigue. Arthritis Rheum. 2007; 56(Suppl 9): 1532.
8. Nashner LM. Practical biomechanics and physiology of balance. In: Jacobson GP, Shepard NT, editors. Balance function assessment and management. 2nd ed. San Diego (CA): Plural Publishing; 2016.
9. Collado-Mateo D, Gallego-Diaz JM, Adsuar JC, Domínguez-Muñoz FJ, Olivares PR, Gusi N. Fear of falling in women with fibromyalgia and its relation with number of falls and balance performance. Biomed Res Int. 2015; 2015: 589014.
10. Key J. 'The core': understanding it, and retraining its dysfunction. J Bodyw Mov Ther. 2013; 17: 541-59.
11. Cabrera-Martos I, Jiménez-Martín AT, López-López L, Rodríguez-Torres J, Ortiz-Rubio A, Valenza MC. Effects of a core stabilization training program on balance ability in persons with Parkinson's disease: a randomized controlled trial. Clin Rehabil. 2020; 34: 764-72.
12. Aggarwal A, Zutshi K, Munjal J, Kumar S, Sharma VP. Comparing stabilization training with balance training in recreationally active individuals. Int J Ther Rehabil. 2010; 17: 244-53.
13. Acar Y, Ilçin N, Gürpinar B, Can G. Core stability and balance in patients with ankylosing spondylitis. Rheumatol Int. 2019; 39: 1389-96.
14. Latikka P, Battie MC, Videman T, Gibbons LE. Correlations of isokinetic and psychophysical back lift and static back extensor endurance tests in men. Clin Biomech (Bristol). 1995; 10: 325-30.
15. Tas NP, Baykara RA, Erdal A, Kal GA. Relationship between balance and vitamin D in fibromyalgia syndrome: a prospective cross-sectional study with HUR BTG4 Balance Master System®. SN Compr Clin Med. 2023; 5: 233.
16. Esfahani NH, Rezaeian ZS, Dommerholt J. The number of repetitions of the McGill tests to reliably determine core muscle endurance in subjects with and without chronic nonspecific low back pain: a cross sectional study. Medical Science. 2019; 23: 452-61.
17. Kim NJ, Yoo KT, An HJ, Shin HJ. The effects of balance exercise on an unstable platform and a stable platform on static balance. J Int Acad Phys Ther Res. 2014; 5: 641-6.
18. Sahin F, Yilmaz F, Ozmaden A, Kotevolu N, Sahin T, Kuran B. Reliability and validity of the Turkish version of the Berg Balance Scale. J Geriatr Phys Ther. 2008; 31: 32-7.
19. Ediz L, Hiz O, Toprak M, Tekeoglu I, Ercan S. The validity and reliability of the Turkish version of the Revised Fibromyalgia Impact Questionnaire. Clin Rheumatol. 2011; 30: 339-46.
20. Garrido-Ardila EM, González-López-Arza MV, Jiménez-Palomares M, García-Nogales A, Rodríguez-Mansilla J. Effectiveness of acupuncture vs. core stability training in balance and functional capacity of women with fibromyalgia: a randomized controlled trial. Clin Rehabil. 2020; 34: 630-45.
21. Rutledge DN, Cherry BJ, Rose DJ, Rakovski C, Jones CJ. Do fall predictors in middle aged and older adults predict fall status in persons 50+ with fibromyalgia? An exploratory study. Res Nurs Health. 2010; 33: 192-206.
22. Kurtais Y, Kutlay S, Ergin S. Exercise and cognitive-behavioural treatment in fibromyalgia syndrome. Curr Pharm Des. 2006; 12: 37-45.

23. Park HY, Jung WS, Kim SW, Lim K. Relationship between sarcopenia, obesity, osteoporosis, and cardiometabolic health conditions and physical activity levels in Korean older adults. Front Physiol. 2021; 12: 706259.

24. Sindwani D, Kaur M. A comparison of core muscle endurance of females with fibromyalgia versus healthy females: an observational study. Rev Colomb Reumatol (Engl Ed). 2024; 31: 438-45.

25. Ambegaonkar JP, Mettinger LM, Caswell SV, Burtt A, Cortes N. Relationships between core endurance, hip strength, and balance in collegiate female athletes. Int J Sports Phys Ther. 2014; 9: 604-16.

26. Gordon AT, Ambegaonkar JP, Caswell SV. Relationships between core strength, hip external rotator muscle strength, and star excursion balance test performance in female lacrosse players. Int J Sports Phys Ther. 2013; 8: 97-104.

27. Kubal S, Kadam V. Relationship between core strength, core endurance and balance in folk dancers - a pilot study. Indian J Physiother Occup Ther. 2021; 15: 157-65.

İstanbul Med J 2026; 27(2): 112-5
DOI: 10.4274/imj.galenos.2026.60598

Urine Beta 2 Microglobulin as a Marker of Recovery in Intrinsic Acute Kidney Injury Patients – a Prospective Observational Study

Ashwin Srinivas, Keshavprakash Viruthagiri, Reena Jose David Xavier, Purnima Ramkumar, Vidya Thirupathipannaiyam Ananthakrishnan

SRM Medical College Hospital and Research Centre, Department of General Medicine, Chengalpattu, India

ABSTRACT

Introduction: Intrinsic acute kidney injury (AKI) is a common and serious syndrome characterized by direct injury to the renal tubules. Although blood creatinine and urea are frequently used to evaluate renal function, they do not sufficiently indicate current tubular damage. Urinary beta 2 microglobulin (β2M) is a lower molecular mass protein reabsorbed via proximal tubules, functions as a possible biomarker for tubular injury and recovery.

Methods: A prospective observational study was carried out over a period of 18 months in the general medicine department of SRM Medical College Hospital and Research Center, enrolling 43 patients over 18 years of age who were diagnosed with AKI. Individuals with chronic kidney disease (CKD) and pre- or post-renal etiologies of AKI were eliminated. Renal parameters such as serum urea, creatinine, urine β2M, electrolytes, complete hemogram, and regular urine analysis, were assessed on admission, 1st, and 3rd months. Ultrasonography was employed to distinguish AKI from CKD. Data were examined to evaluate changes in renal variables over time.

Results: On admission, all patients exhibited elevated levels of urea, creatinine, and urinary β2M. At three months, serum urea and creatinine levels normalized in all patients, although 77.8% still exhibited elevated β2M levels. Notable mean decreases were recorded in urea, creatinine, and β2M of 13.22 mg/dL, 0.74 mg/dL and 11.2 µg/L respectively from starting period to three months. No statistically significant correlation was identified between gender and renal parameters.

Conclusion: Urinary β2M serves as a significant biomarker for identifying ongoing tubular injury in intrinsic AKI, despite the normalization of standard renal indicators.

Keywords: Acute kidney injury, tubular damage, urine β2 microglobulin, renal function

Introduction

Acute kidney injury (AKI) is a term used to describe a heterogeneous group of illnesses that share similar diagnostic characteristics, such as an increase in serum creatinine (SCr) concentration, frequently accompanied by a decrease in urine volume. Up to 30% of intensive care unit (ICU) admissions and 5–7% of acute-care hospital admissions are complicated by AKI (1). The severity of AKI can vary from mild and temporary alterations in glomerular filtration rate laboratory parameters to severe and quickly fatal disruptions in the kidney's capacity to maintain proper circulating volume regulation, eliminate nitrogenous wastes and metabolic toxins, and preserve the electrolyte and acid-base composition of plasma (2).

Because of several factors—including the impact of muscle mass, fluid status, or a delayed rise in creatinine level following the onset of renal injury—SCr is a poor diagnostic marker for AKI, making early intervention difficult (3). Several new biomarkers outperformed SCr in the diagnosis of AKI, and many were strongly correlated with immediate diagnostic results. When tubular damage or inflammation occurs, the majority of biomarkers are increased and stay that way until the injury or inflammation goes away (4).

Numerous proteins, including enzymes and urine beta 2 microglobulin (β2M), are filtered by glomeruli before being reabsorbed in the proximal tubule (5,6). β2M begins to rise early in renal failure and is unaffected by muscle mass. Reduced reabsorption of β2M and tubular enzymes as a result of tubular damage raises urine concentrations (7). These indicators might be useful markers of fibrosis or tubular damage. Serum β2M has been suggested as a potential measure to evaluate kidney function in AKI and chronic kidney disease (CKD) due to these characteristics (8).

Among existing tubular biomarkers, urinary β2M offers several practical advantages over alternatives such as neutrophil gelatinase-associated

Address for Correspondence: Prof. Vidya Thirupathipannaiyam Ananthakrishnan, MD, SRM Medical College Hospital and Research Centre, Department of General Medicine, Chengalpattu, India
E-mail: vidyaa@srmist.edu.in **ORCID ID:** orcid.org/0000-0002-8717-7866

Cite this article as: Srinivas A, Viruthagiri K, David Xavier RJ, Ramkumar P, Thirupathipannaiyam Ananthakrishnan V. Urine beta 2 microglobulin as a marker of recovery in intrinsic acute kidney injury patients – a prospective observational study. İstanbul Med J. 2026; 27(2): 112-5

Received: 23.09.2025
Accepted: 08.03.2026
Publication Date: 12.05.2026

lipocalin, kidney injury molecule-1, and cystatin C. Urinary β2M is a low-molecular-weight protein (11.8 kDa) that is freely filtered at the glomerulus and nearly entirely reabsorbed by proximal tubular cells under normal circumstances; therefore, its rise in urine particularly indicates compromised tubular reabsorptive ability (9,10). Moreover, β2M tests are readily accessible in standard clinical laboratories, necessitate no specialised apparatus, and are significantly more economical than multiplex biomarker panels (11). These attributes render urine β2M a practical and dependable option for assessing tubular integrity in resource-constrained clinical environments.

Methods

Study Design

This prospective observational study was conducted in the Department of General Medicine at SRM Medical College. Over the course of 18 months, 43 patients participated in this study. The study population was calculated using the given formulas. This study included individuals over the age of 18 who were diagnosed with AKI and sepsis, and who were taking nephrotoxic medications. Individuals receiving chronic renal replacement treatment and those who had pre- and post-renal acute kidney damage (CKD) were not included.

Investigations and Follow-Up

Renal function tests, such as serum urea and creatinine, β2M, urine routine analysis, serum electrolytes, and complete hemogram, were measured at the time of admission. Ultrasonography of the abdomen was performed to differentiate AKI from CKD. These parameters were measured at 1 and 3 months.

Ethics and Consent

The study protocol was approved by the SRM Medical College Hospital and Research Centre Institutional Ethics Committee (registration number: IEC-ST1023-1717, date: 06.11.2023). Written informed consent was obtained from all participants included in the study. The participants were informed about the nature, purpose, and procedures of the study, and their right to withdraw at any time without any consequences. Confidentiality and anonymity of the participants were strictly maintained throughout the study.

Defining Outcomes

AKI Defined by KDIGO

- Increase in SCr by more than or equal to 0.3 mg/dL within 48 hours
- Increase in SCr to more than or equal to 1.55 to 1.9 times baseline, known or presumed to have occurred in the past 7 days
- Urine volume <0.5 mL/kg/hr for 6 hours

Statistical Analysis

The data were examined with SPSS version 22. Continuous variables were expressed as mean ± standard deviation. Renal function markers (urea, creatinine, and β2M), assessed at admission, one month, and three months, were analysed using repeated-measures analysis of variance. A p value less than 0.05 was deemed statistically significant.

Results

The average urea levels considerably declined from admission to three months (48.63±1.3 to 35.42±0.9 mg/dL). Despite a decrease in creatinine (1.86±0.2 to 1.11±0.5 mg/dL) and β2M (388.85±1.7 to 377.65±1.4 µg/L) levels, the alterations were not statistically significant. However, renal function improved during the follow-up period (Table 1).

All patients had abnormal urea, creatinine, and β2M levels when they were evaluated at admission (0 months). After a month of follow-up, 82.2%, 64.4%, and 88.9% of the AKI patients had abnormal urea, creatinine, and β2M readings, respectively. However, after three months' follow-up, all patients had urea and creatinine levels within acceptable ranges. β2M was abnormal in 77.8% of cases (Figures 1-3).

Table 2 displays the analysis of mean renal parameter values over the time frame. The average difference in levels of urea over three months was 13.2±2.3 mg/dL, also indicating a statistically significant decrease. Creatinine levels diminished by 0.74±0.03 mg/dL; however, this alteration was not considered statistically significant. β2M levels demonstrated a notable decrease of 11.2±2.0 µg/L over a period of 3 months.

Table 1. Renal function parameters

Renal function values (mean)	At admission	1 month	3 month	p value
Urea (mg/dL)	48.63±1.3	43.5±2.5	35.42±0.9	0.03
Creatinine (mg/dL)	1.86±0.2	1.57±0.3	1.11±0.5	0.2
β2M (µg/L)	388.85±1.7	382.91±2.8	377.65±1.4	0.4

β2M: Beta 2 microglobulin

Table 2. Mean differences of different parameters

Parameters	Mean difference			p value
	0-1 month	1-3 months	0-3 months	
Urea	5.13±0.9	8.1±1.2	13.2±2.3	0.002
Creatinine	0.28±0.03	0.46±0.06	0.74±0.03	0.1
β2M	5.94±0.8	5.26±1.1	11.2±2.0	0.03

β2M: Beta 2 microglobulin

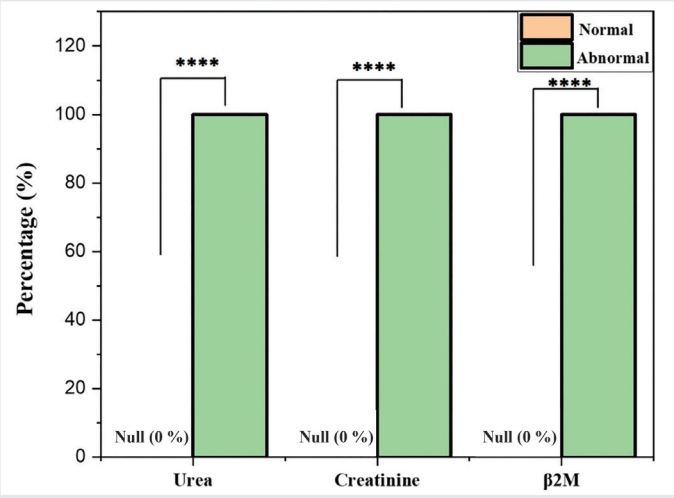

Figure 1. Renal values at admission. **** indicates the renal values are extremly significant

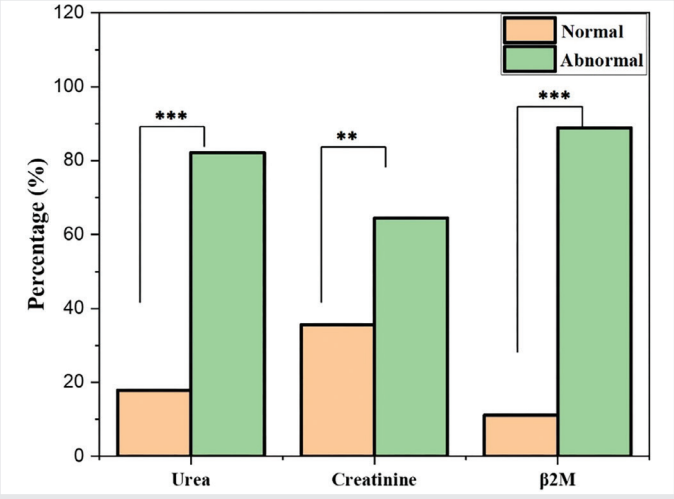

Figure 2. Renal values after 1 month of follow up. Values marked with asterisks indicate levels of statistical significance: **: indicates moderately significant differences, and *** indicates highly statistically significant differences in renal values

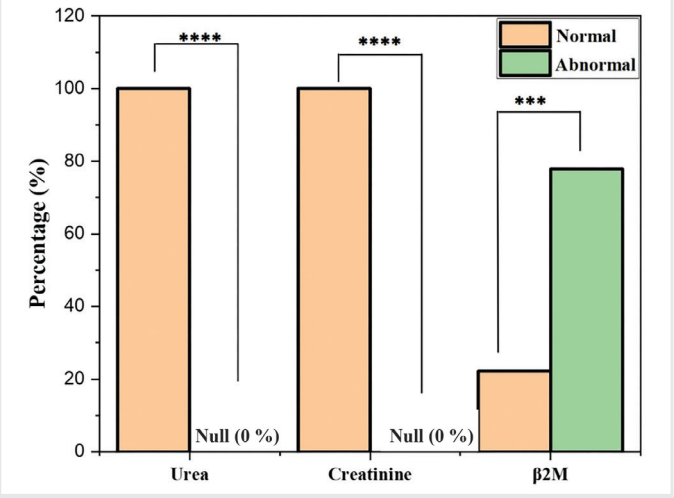

Figure 3. Renal values after 3 months of follow-up. Values marked with asterisks indicate levels of statistical significance: *** indicates highly significant differences, and **** indicates extremely significant differences in renal values

Discussion

In this study, patients with AKI exhibited a gradual enhancement in renal function indicators over a 3-month follow-up period. The study findings align with other finding, who documented substantial enhancements in renal function subsequent to proper intervention in intrinsic AKI (12).

SCr levels exhibited a reduction from period of admission to three months, although the alteration was not statistically significant. The average difference over 0-3 months was 0.74±0.03 mg/dL, indicating a progressive enhancement in glomerular filtration performance. Chapman et al. (13) observed analogous results, indicating that SCr may normalize at a slower rate and is affected by variables such as muscle mass and hydration condition, hence constraining its sensitivity to early tubular injury.

Conversely, β2M exhibited a more gradual decrease from admission to three months. The average difference during 0-3 months was 11.2±2.0 µg/L. This signifies that, although glomerular function improved, tubular recovery remained delayed. The results align with those of Argyropoulos et al. (7), who highlighted β2M as a sensitive indicator of ongoing tubular dysfunction despite the normalization of SCr levels.

The sustained increase in β2M after three months, as demonstrated in our study, underscores the continuous tubular damage that may elude detection by standard markers. Other researches claimed that enduring tubular dysfunction following AKI may facilitate the progression to CKD, despite corrected serum urea and creatinine levels (13-15).

Numerous investigations have confirmed β2M as an early prognostic biomarker for the severity and course of AKI. Shahjahan et al. (16) indicated that β2M provides enhanced sensitivity relative to creatinine for identifying renal impairment at multiple stages.

Our data indicate that whereas glomerular markers such as urea and creatinine may return to normal within three months post-AKI, tubular markers like β2M could stay at elevated levels, signifying subclinical injury and an increased CKD risk progression. This highlights the need to integrate tubular biomarkers into standard monitoring regimens for patients with AKI, enabling earlier management and preservation of long-term renal function.

Current research indicates that urine β2M levels over 300 µg/L may signify considerable proximal tubular dysfunction, necessitating further clinical monitoring. In the setting of AKI-to-CKD progression, consistently elevated β2M levels—especially those over 200 µg/L beyond 90 days post-AKI—are linked to inadequate tubular healing and an increased risk of long-term renal decline (17-19). Subsequent prospective research involving larger cohorts should validate specific cut-off values across diverse AKI patient populations to establish standardised clinical criteria for risk classification.

Study Limitations

The current study has a limited sample size, perhaps constraining the generalisability of the results. Although the sample size was determined to satisfy the minimum statistical criteria for this single-centre observational study, a larger multi-centre cohort would enhance the generalisability and statistical power of the results and is recommended

for future investigations. The follow-up duration was restricted to three months, precluding the assessment of long-term outcomes such as the progression to chronic renal disease. Urine β2M levels are influenced by urine pH, collection techniques, and storage conditions, which were not entirely standardised. Furthermore, since this was a single-center observational study, the findings may have limited generalisability to other demographic groups or clinical settings.

Conclusion

This study demonstrated a considerable improvement in urea and creatinine levels within three months post-AKI, indicating successful glomerular recovery. Nonetheless, sustained elevation of β2-microglobulin indicates continued tubular damage despite adjusted conventional indicators. β2M may function as a sensitive marker for assessing subclinical renal impairment. The integration of tubular biomarkers may improve the long-term care and prognosis of patients with AKI.

Ethics

Ethics Committee Approval: The study protocol was approved by the SRM Medical College Hospital and Research Centre Institutional Ethics Committee (registration number: IEC-ST1023-1717, date: 06.11.2023).

Informed Consent: Written informed consent was obtained from all participants included in the study. The participants were informed about the nature, purpose, and procedures of the study, and their right to withdraw at any time without any consequences. Confidentiality and anonymity of the participants were strictly maintained throughout the study.

Acknowledgments

We would like to thank SRM Medical College Hospital and Research Centre, Faculty of Medicine and Health Sciences, SRMIST, Kattankulathur, for their cordial support of this study. We also extend our gratitude to Dr. Vishnupriya Subramaniyan, Ph.D., Research Writer, SRM Medical College Hospital and Research Centre, SRMIST, for her valuable assistance in the preparation of this manuscript.

Footnotes

Authorship Contributions: Surgical and Medical Practices - A.S., R.J.D.X., P.R.; Concept - A.S., K.V., V.T.A.; Design - R.J.D.X., P.R., V.T.A.; Data Collection or Processing - A.S., K.V., R.J.D.X., P.R., V.T.A.; Analysis or Interpretation - A.S., R.J.D.X., P.R., V.T.A.; Literature Search - A.S., K.V., R.J.D.X., V.T.A.; Writing - A.S., P.R., V.T.A.

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. Makris K, Spanou L. Acute kidney injury: definition, pathophysiology and clinical phenotypes. Clin Biochem Rev. 2016; 37: 85-98.
2. Goyal A, Daneshpajouhnejad P, Hashmi MF, Bashir K. Acute kidney injury. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2025 Jan.
3. Kellum JA, Romagnani P, Ashuntantang G, Ronco C, Zarbock A, Anders HJ. Acute kidney injury. Nat Rev Dis Primers. 2021; 7: 52.
4. Rossiter A, La A, Koyner JL, Forni LG. New biomarkers in acute kidney injury. Crit Rev Clin Lab Sci. 2024; 61: 23-44.
5. Molitoris BA, Sandoval RM, Yadav SPS, Wagner MC. Albumin uptake and processing by the proximal tubule: physiological, pathological, and therapeutic implications. Physiol Rev. 2022; 102: 1625-67.
6. Fels J, Scharner B, Zarbock R, Zavala Guevara IP, Lee WK, Barbier OC, et al. Cadmium complexed with β2-microglubulin, albumin and lipocalin-2 rather than metallothionein cause megalin:cubilin dependent toxicity of the renal proximal tubule. Int J Mol Sci. 2019; 20: 2379.
7. Argyropoulos CP, Chen SS, Ng YH, Roumelioti ME, Shaffi K, Singh PP, et al. Rediscovering beta-2 microglobulin as a biomarker across the spectrum of kidney diseases. Front Med (Lausanne). 2017; 4: 73.
8. Puthiyottil D, Priyamvada PS, Kumar MN, Chellappan A, Zachariah B, Parameswaran S. Role of urinary beta 2 microglobulin and kidney injury molecule-1 in predicting kidney function at one year following acute kidney injury. Int J Nephrol Renovasc Dis. 2021; 14: 225-34.
9. Zhang B, Li Q. Urine. In: Clin Mol Diagn. Singapore: Springer Singapore; 2021. p. 241-52.
10. Wong J, Kaja Kamal RM, Vilar E, Farrington K. Measuring residual renal function in hemodialysis patients without urine collection. Semin Dial. 2017; 30: 39-49.
11. Nair P, Bindhu OS, Mussavira S. Biomarkers. In: Approved: The Life Cycle of Drug Development. Cham: Springer Nature Switzerland; 2025. p. 333-63.
12. James MT, Bhatt M, Pannu N, Tonelli M. Long-term outcomes of acute kidney injury and strategies for improved care. Nat Rev Nephrol. 2020; 16: 193-205.
13. Chapman CL, Holt SM, O'Connell CT, Brazelton SC, Howells WAB, Medved HN, et al. Acute kidney injury biomarkers and hydration assessments following prolonged mild hypohydration in healthy young adults. Am J Physiol Renal Physiol. 2023; 325: F199-213.
14. Wang Y, Hu L, Zhou D, Chen X, Zhou L. Association of urinary biomarkers of renal tubular injury with cognitive dysfunction in older patients with chronic kidney disease: a cross-sectional observational study. Brain Sci. 2023; 13: 551.
15. Chawla LS, Amdur RL, Shaw AD, Faselis C, Palant CE, Kimmel PL. Association between AKI and long-term renal and cardiovascular outcomes in United States veterans. Clin J Am Soc Nephrol. 2014; 9: 448-56.
16. Shahjahan, Yasmin R, Mahsud MAJ, Khan M, Subzwari J, Hussain J. Correlation of beta-2 microglobulin with serum creatinine and creatinine clearance in patients with different levels of renal function. Gomal J Med Sci. 2011; 9: 178-82.
17. Jaswanth C, Priyamvada PS, Zachariah B, Haridasan S, Parameswaran S, Swaminathan RP. Short-term changes in urine beta 2 microglobulin following recovery of acute kidney injury resulting from snake envenomation. Kidney Int Rep. 2019; 4: 667-73.
18. Zeng X, Hossain D, Bostwick DG, Herrera GA, Zhang PL. Urinary β2-microglobulin is a good indicator of proximal tubule injury: a correlative study with renal biopsies. J Biomark. 2014; 2014: 492838.
19. Wen Y, Xu L, Melchinger I, Thiessen-Philbrook H, Moledina DG, Coca SG, et al. Longitudinal biomarkers and kidney disease progression after acute kidney injury. JCI Insight. 2023; 8(9): e167731.

İstanbul Med J 2026; 27(2): 116-21
DOI: 10.4274/imj.galenos.2026.23723

Evaluation of Organ Doses to Female Reproductive System During Abdominal CT Imaging: A Phantom Study

Duygu Tunçman Kayaokay[1], Berrin Yalçın[2], Aysun Özsoy Ata[2], Özge Coşkun Sağlam[3], Osman Günay[4], Mustafa Demir[5], Fahrettin Fatih Kesmezacar[6]

[1]İstanbul University–Cerrahpaşa, Vocational School of Health Services, Radiotherapy Program, İstanbul, Türkiye
[2]University of Health Sciences Türkiye, İstanbul Training and Research Hospital, Clinic of Radiation Oncology, İstanbul, Türkiye
[3]İstanbul Bilgi University Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, İstanbul, Türkiye
[4]Yıldız Technical University Faculty of Electrical and Electronics, Department of Biomedical Engineering, İstanbul, Türkiye
[5]İstanbul University–Cerrahpaşa, Cerrahpaşa Faculty of Medicine, Department of Nuclear Medicine, İstanbul, Türkiye
[6]İstanbul University–Cerrahpaşa, Vocational School of Health Services, Medical Imaging Techniques, İstanbul, Türkiye

ABSTRACT

Introduction: This study aimed to determine the absorbed radiation doses in radiosensitive female pelvic organs during abdominal computed tomography (CT) imaging using thermoluminescent dosimeters (TLD-100) and an anthropomorphic phantom.

Methods: A Computerized Imaging Reference Systems, Inc. female pelvic anthropomorphic phantom was used to simulate realistic human anatomy. TLD-100 dosimeters were positioned within organ-equivalent cavities corresponding to the ovaries, uterus, and urinary bladder. CT scans were performed using a Toshiba Aquilion 64-slice scanner with standard abdominal parameters (120 kVp, 300 mA, 5 mm slice thickness, pitch 1.0). After exposure, TLDs were read using a Harshaw 3500 reader, and organ doses were calculated based on calibration coefficients.

Results: The urinary bladder exhibited the highest mean absorbed dose (27.89 mGy), followed by the fundus uteri (26.34 mGy), left ovary (23.22 mGy), cervix uteri (21.88 mGy), and right ovary (20.91 mGy). Lower doses were measured in deeper or more peripheral regions, such as the recessus rectouterina (19.37 mGy) and the medulla spinalis (16.73 mGy).

Conclusion: The results indicate that even under standard clinical protocols, radiosensitive organs of the female reproductive system receive measurable radiation doses during abdominal CT. These findings emphasize the need for protocol optimization and shielding strategies to minimize exposure, especially in reproductive-age women.

Keywords: Experimental dosimetry, phantom study, female reproductive system

Introduction

One of the most popular imaging modalities in diagnostic radiology is computed tomography (CT), which produces high-resolution cross-sectional pictures that enable in-depth analysis of interior organs and anatomical structures (1). It is essential for the evaluation of abdominal and pelvic disorders, including neoplastic, inflammatory, and traumatic conditions, because of its speed, diagnostic precision, and accessibility. The evaluation of gynecologic tumors, staging of pelvic malignancies, identification of metastatic illness, monitoring of treatment response, and differential diagnosis of acute abdominal discomfort are among the various uses of CT of the female reproductive system in clinical practice (2,3). Additionally, CT can be used in emergency settings when magnetic resonance imaging (MRI) or ultrasound are unavailable or inconclusive (4). Nevertheless, CT produces comparatively larger radiation doses than other imaging methods despite its clinical benefits, making up between 60 and 70 percent of all medical radiation exposure globally (5). The rising frequency of CT scans, particularly in women of reproductive age, has raised concerns about unnecessary radiation exposure to radiosensitive organs such as the ovaries, uterus, and bladder. Optimizing imaging procedures and guaranteeing patient safety in such situations require an understanding of the radiation doses specific to each organ. Ionizing radiation is especially dangerous to the female reproductive system, which includes radiosensitive organs such as the uterus, ovaries, and bladder. Even though they are not the imaging objective, these organs are frequently found inside or close to the primary beam during abdominal and pelvic CT exams, resulting in quantifiable radiation exposure (6). Because the ovaries have a limited supply of germ cells, exposure to ionizing radiation can result in temporary or permanent

Address for Correspondence: Asst. Prof. Fahrettin Fatih Kesmezacar, MD, İstanbul University–Cerrahpaşa, Vocational School of Health Services, Medical Imaging Techniques, İstanbul, Türkiye
E-mail: f.kesmezacar@iuc.edu.tr **ORCID ID:** orcid.org/0000-0001-5110-1184

Cite this article as: Tunçman Kayaokay D, Yalçın B, Özsoy Ata A, Coşkun Sağlam Ö, Günay O, Demir M, et al. Evaluation of organ doses to female reproductive system during abdominal CT imaging: a phantom study. İstanbul Med J. 2026; 27(2): 116-21

Received: 02.12.2025
Accepted: 24.03.2026
Publication Date: 12.05.2026

infertility, depending on the patient's age and the radiation dose. The uterus is composed of proliferative, vascularized tissues. Reproductive outcomes may be affected by predictable effects, including fibrosis and endometrial shrinkage. Despite not being a reproductive organ, the bladder is physically close to the pelvic structures and receives a comparatively higher dose because of its central location and close proximity to the scan field (7,8).

Deterministic or stochastic biological consequences, such as decreased fertility, teratogenic results in women of reproductive age, and an elevated lifetime risk of radiation-induced cancers in radiosensitive pelvic tissues, can arise from even low to moderate radiation doses (9). Precise measurement of organ-specific absorbed doses in abdominal CT imaging is crucial to assess potential hazards and optimize radiation protection procedures for female patients.

Using anthropomorphic phantoms and a range of dosimetric devices, including thermoluminescent dosimeters (TLDs), optically stimulated luminescence dosimeters (OSLDs), and metal oxide semiconductor field-effect transistors (MOSFETs), a number of studies have recently examined organ-specific dose distributions during CT imaging (10,11). Regarding sensitivity, spatial resolution, and reusability, each system offers unique benefits. Because of its near-tissue equivalency, high sensitivity to low photon energies, linear dose-response characteristics, and capacity to measure cumulative absorbed dose over multiple exposures, TLD-100 (LiF:Mg,Ti) continues to be one of the most dependable and extensively used detectors in medical physics research. TLDs are especially well-suited for dose measurement in CT applications because they show no energy dependency within the diagnostic X-ray range (12).

The Computerized Imaging Reference Systems (CIRS), Inc. ATOM® female model and other anthropomorphic phantoms are essential for measuring radiation exposure in practical healthcare settings. Dosimeters may be precisely localized within organ-equivalent cavities or slices using these phantoms, which are constructed from tissue-equivalent materials that replicate the radiological characteristics of human organs. Anthropomorphic phantoms, in contrast to basic geometric or aqueous phantoms, mimic the intricate anatomical interactions among organs and account for variations in attenuation, dispersion, and absorption across different tissues. These setups preserve repeatability and control over imaging settings while enabling precise calculation of organ doses. Additionally, multiple dosimetric studies have demonstrated that the combination of anthropomorphic phantoms and TLD-100 dosimetry yields dose measurements that closely correlate with Monte Carlo simulation results and patient-derived dosimetric data, thereby validating their reliability for experimental radiation dose assessment in diagnostic radiology.

To measure the radiation doses absorbed by vital organs of the female reproductive system, this study used a CIRS female anthropomorphic phantom to simulate a realistic clinical scenario for abdominal CT imaging. We aimed to experimentally replicate the dose distribution that occurs during standard diagnostic CT procedures by positioning TLD-100 dosimeters in organ-equivalent regions of the phantom. This could provide information to help optimize dosing and protective measures for female patients.

Despite the increasing number of CT examinations, limited experimental dosimetric data that specifically quantify radiation doses to individual female pelvic organs using anthropomorphic phantoms with precisely placed detectors. Addressing this gap, the present study aims to experimentally evaluate organ-specific absorbed doses to radiosensitive female pelvic organs during abdominal CT imaging, using TLD-100 dosimeters positioned within a CIRS anthropomorphic female phantom. By replicating a realistic clinical imaging scenario, this study provides experimentally derived dose data that may contribute to improved radiation protection and dose optimization strategies for female patients undergoing CT examinations.

Based on this framework, we hypothesized that radiosensitive female pelvic organs receive measurable and clinically relevant radiation doses during routine abdominal CT examinations, and that these doses can be quantified experimentally using anthropomorphic phantom-based TLD-100 dosimetry.

Methods

Ethical approval for this study was obtained from a certified University of Health Sciences Türkiye, İstanbul Training and Research Hospital Clinical Research Ethics Committee (approval number: 281, date: 07.11.2025). As this study involved only phantom measurements and did not include human participants, informed consent was not required.

In this experimental dosimetric work, the human pelvic anatomy was simulated for dose quantification in abdominal CT using a CIRS anthropomorphic female pelvic phantom (Norfolk, Virginia, USA) (13).

As shown in Figure 1, the phantom represents the female lower abdomen and pelvic region and consists of axial slices that mimic the radiological features of the human body. Each slice is approximately 2.5 cm thick and is labeled numerically from superior to inferior (sections 1–16 anteriorly and 28–38 posteriorly), allowing reproducible alignment during assembly and consistent dosimeter placement across measurements.

The phantom enables realistic simulation of X-ray attenuation and scatter during CT imaging by replicating major pelvic structures such as the bladder, uterus, ovaries, rectum, and the surrounding soft tissues. Because of its modular architecture, dosimeters could be inserted at certain anatomical levels corresponding to clinical imaging planes and in organ-equivalent cavities. Exterior reference lines and numbering visible on the phantom's anterior and posterior surfaces facilitated precise slice identification and repositioning between scans.

TLDs (TLD-100, LiF:Mg,Ti), which are frequently used for diagnostic X-ray dosimetry due to near-tissue equivalency, high sensitivity, and repeatability, were utilized to assess radiation doses. To remove any residual signals, all TLD chips were annealed in a nitrogen atmosphere at 400 °C for one hour and at 100 °C for two hours prior to exposure. A reference X-ray beam (120 kVp, 10 mm Al filtration), traceable to a secondary standards dosimetry facility, was used to calibrate the dosimeters. Calibration procedures were performed at a national Secondary Standard Dosimetry Laboratory accredited for diagnostic X-ray dosimetry. The measured charge values were then converted into absorbed doses using the calibration coefficients (mGy/nC) obtained.

The overall measurement uncertainty associated with the TLD system, including calibration and reader-related factors, is typically reported to lie within 5–10% for diagnostic X-ray energy ranges.

TLD-100 dosimeters were positioned in organ-equivalent areas of the CIRS female pelvic anthropomorphic phantom to assess absorbed doses during abdominal CT imaging, as illustrated in Figure 2 and described in Table 1. The bladder, uterus, bilateral uterine tubes, right and left ovaries, rectum, and soft-tissue backdrop areas were all designated as cavities into which dosimeters were placed. To provide precise anatomical localization, each cavity was carefully chosen based on distinguishable bony features seen on the appropriate axial slices of the phantom.

For anatomical orientation, the uterine tubes extend laterally from the uterine cornua within the true pelvis, while the ovaries are located on the lateral pelvic wall within the ovarian fossa (14-17). The uterus is positioned centrally between the bladder anteriorly and the rectum posteriorly, reflecting its normal anatomical relationship (17,18). This anatomical framework facilitated accurate placement of TLD-100 detectors within organ-equivalent regions of the phantom.

Figure 1. Anterior (left) and posterior (right) views of the CIRS anthropomorphic female pelvic phantom used in this study. The numbered axial sections allow accurate localization and reproducible placement of TLD-100 dosimeters within organ-equivalent cavities for radiation dose measurements during abdominal CT imaging. CIRS: Computerized Imaging Reference System, CT: Computed tomography

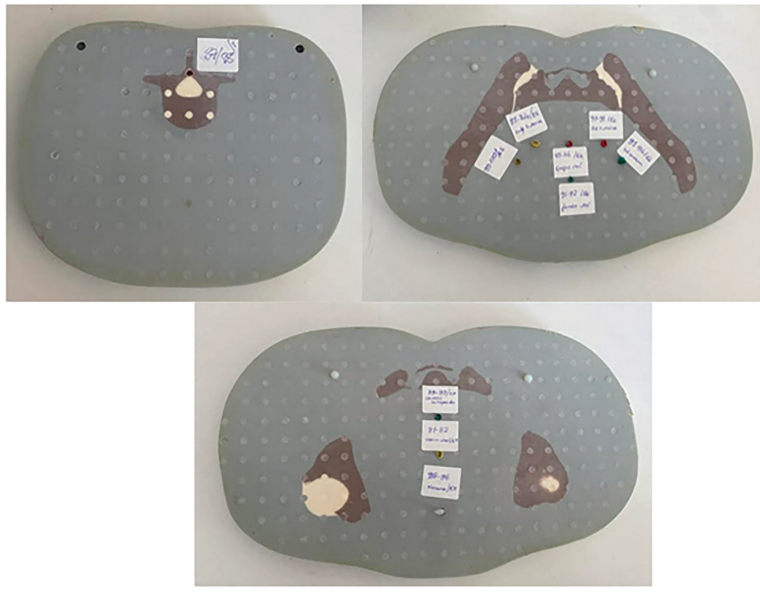

Figure 2. Dosimeters were inserted into cavities corresponding to the ovaries, uterus, and urinary bladder to measure absorbed doses during abdominal CT examinations. CT: Computed tomography

Table 1. Placement of TLD-100 dosimeters in organ-equivalent regions of the CIRS female pelvic phantom

Slice	Organ	TLD Number	TLD count per organ
1	Spinal cord	87–88	2
6	Right ovary	99–100	2
6	Left ovary	93–94	2
6	Uterine tube-right	83–84	2
6	Uterine tube-left	97–98	2
6	Corpus uteri	85–86	2
6	Fundus uteri	91–92	2
7	Bladder	95–96	2
7	Cervix uteri	81–82	2
7	Recessus rectouterina	89–90	2

A Toshiba Aquilion 64-slice CT scanner (Japan) was used to acquire CT images. To guarantee precise reproducibility, the CIRS female pelvic anthropomorphic phantom was placed supine on the scanner table and oriented at the isocenter using the integrated laser positioning system (see Figure 3). To replicate clinical imaging protocols commonly used in female patients, a standard abdominopelvic CT procedure was chosen.

The scanning parameters were: tube voltage of 120 kVp, tube current of 300 mA, rotation duration of 0.75 s, pitch of 1.0, and slice thickness of 2 mm. The entire pelvic region was covered by the scan length, which ran from the level of the iliac crest to the inferior boundary of the ischial tuberosities. Using the helical scan mode (GG-Hel), the field of view was adjusted to 552.3 mm (XL mode). The overall scan time was approximately 21 seconds, and the effective mAs was 240. To maintain consistent parameters during the process, automatic exposure control (AEC) was turned off. For optimal soft-tissue contrast visualization, the reconstruction matrix was set to 512 × 512 pixels, and the window level and window width were adjusted to 90 and 140, respectively. The total dose-length product for the acquisition was approximately 1143.7 mGy·cm, and the CT dose index volume was 26.9 mGy, as recorded on the scanner console.

Statistical Analysis

Because the study was limited to TLD placement on the CIRS anthropomorphic female pelvic phantom and direct dose readouts, no statistical analysis was required.

Results

The phantom image, which confirms correct positioning of the phantom within the CT scanner and full coverage of the targeted pelvic anatomy by the scanning area, provides crucial validation of the experimental setup (see Figure 4). The phantom's tissue-equivalent composition faithfully mimics human anatomy, as evidenced by the clear difference between bone and soft-tissue densities. An internal validation of the anatomical precision of TLD placement was achieved by visualizing specific pelvic organs on the obtained CT images. As a result, the radiation doses measured in organ-equivalent regions using TLD-100 dosimeters can be regarded as representative of those in actual clinical settings.

Following CT imaging, absorbed doses measured by TLD-100 dosimeters placed in the organ regions of interest of the CIRS female pelvic phantom were examined. Table 2 displays the measured mean organ doses (mGy) for each anatomical structure.

Discussion

The study objectively evaluated, using an anthropomorphic female phantom and a TLD-100 dosimeter, the radiation doses absorbed by radiosensitive female pelvic organs during abdominal CT scans. According to experimental results, the uterine fundus (26.34 mGy), the left ovary (23.22 mGy), and the bladder (27.89 mGy) received the highest

Figure 3. Positioning of the CIRS female pelvic anthropomorphic phantom on the Toshiba Aquilion 64-slice CT scanner. The phantom was placed in the supine position and aligned at the scanner isocenter using the built-in laser positioning system to ensure reproducible imaging geometry during abdominal–pelvic CT acquisitions. CIRS: Computerized Imaging Reference System, CT: Computed tomography

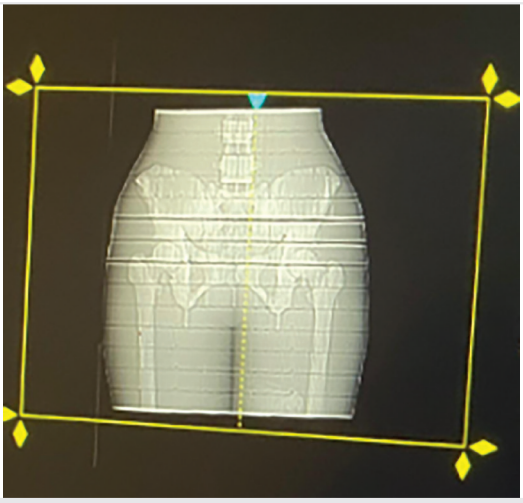

Figure 4. Computed tomography image of the CIRS female pelvic anthropomorphic phantom acquired with the Toshiba Aquilion 64-slice scanner. CIRS: Computerized Imaging Reference Systems

Table 2. Mean absorbed doses measured by TLD-100 dosimeters in female pelvic organs during abdominal CT imaging

Organ	Dose (mean mGy)
Spinal cord	16.73
Right ovary	20.91
Left ovary	23.22
Uterine tube-right	20.75
Uterine tube-left	21.57
Corpus uteri	21.35
Fundus uteri	26.34
Bladder	27.89
Cervix uteri	21.88
Recessus rectouterina	19.37

TLD: Thermoluminescent dosimeter, CT: Computed tomography

mean absorbed doses. These findings align with the anatomical position of the pelvic structures with respect to the CT beam path.

During abdominal CT scans, the bladder, located centrally in the pelvis, occupies the anterior region directly aligned with the primary X-ray beam. The bladder's higher absorbed dose, compared with surrounding organs, can be explained by its relatively large volume and proximity to the beam isocenter. Because the fundal area of the uterus is located above the bladder, it is particularly vulnerable to both primary and scattered radiation. The ovaries, situated laterally, close to the iliac fossae, were exposed to moderate doses (20.91-23.22 mGy), suggesting that the surrounding soft tissues were partially spared, but dispersed radiation exposure persisted.

The range of variability documented in earlier clinical research is consistent with our measured ovarian and uterine dose levels. Depending on scanner technology, Obed et al. (19) reported uterine doses of 12–43 mGy and ovarian doses of 11–33 mGy. Bladder and gonadal doses of 13.6±1.9 mGy and 13.0±1.9 mGy, respectively, were reported by Gao et al. (20). The relatively higher values observed in the present study may be related to differences between phantom-based experimental conditions and patient-based measurements, including variations in tissue composition and attenuation characteristics.

Furthermore, our study provides direct organ-specific dose measurements obtained by anatomically accurate placement of TLD-100 detectors within an anthropomorphic phantom, thereby addressing limitations highlighted in previous studies such as the lack of phantom-based validation reported by Obed et al. (19). In a recent large-scale patient-based investigation, Shubayr and Alashban (21) analyzed 665 abdomen–pelvis CT examinations and reported a mean uterine dose of 10.86±6.09 mGy (range: 2.13–24.06 mGy). These values are lower than the doses observed in our phantom measurements (21.35–26.34 mGy for the uterus), a difference that may reflect clinical variability in patient anatomy, scanning protocols, and physiological conditions. Phantom-based and patient-based studies together provide complementary information for understanding radiation exposure of female reproductive organs during pelvic CT imaging.

The rectouterine recessus (Douglas's pouch), a posterior pelvic anatomical landmark, was incorporated into the study to further assess the spatial dosage gradient. This region's significantly lower absorbed radiation (19.37 mGy) supports the intrinsic attenuation resulting from deeper soft tissue and sacral shielding. This measurement is crucial for radiation protection because it facilitates accurate scan length optimization and avoids needless exposure of posterior pelvic structures, such as bowel loops.

From a clinical standpoint, this study emphasizes the importance of optimization techniques, such as AEC, iterative reconstruction, and scan-interval minimization to reduce exposure while preserving diagnostic quality, because even standard abdominal CT protocols deliver specific doses to the reproductive organs.

Study Limitations

This study has several limitations. First, the measurements were performed using a single CT scanner and acquisition protocol, which may limit the generalizability to other CT systems or clinical protocols. Second, the study was conducted using an anthropomorphic phantom rather than on real patients. Although such phantoms are widely used in experimental dosimetry studies and provide controlled and reproducible measurement conditions, they cannot fully represent patient-specific anatomical variability and tissue heterogeneity.

Another limitation is that only TLD-100 dosimeters were used for dose measurements. Comparisons with other dosimetric techniques such as OSLDs, MOSFET detectors, or Monte Carlo–based dose simulations were not included. Future studies should incorporate multiple scanners, different CT protocols, advanced dosimetric techniques, and computational simulations to further validate and extend the present findings.

Conclusion

Using a realistic anthropomorphic female phantom and TLD-100 dosimetry, this study showed that radiosensitive female pelvic organs receive clinically relevant radiation doses during abdominal CT imaging, with a distinct dose gradient varying with anatomical position. The

following basic radiation protection principles corroborate the study's findings:

Justification: Ionizing radiation scans (such as CT) should only be performed when clinically indicated; whenever possible, alternatives such as ultrasonography or MRI should be considered.

Optimization (as low as reasonably achievable): Using dose-modulated technology, clinicians should adjust the scan parameters to the patient's anatomy and diagnostic requirements.

Dose awareness: Patient safety depends on ongoing dose index monitoring and adherence to DRLs.

These results emphasize the significance of protecting the reproductive organs in clinical female pelvic CT applications and offer insights into patient-centered dose optimization techniques.

Ethics

Ethics Committee Approval: Ethical approval for this study was obtained from a certified University of Health Sciences Türkiye, İstanbul Training and Research Hospital Clinical Research Ethics Committee (approval number: 281, date: 07.11.2025).

Informed Consent: As this study involved only phantom measurements and did not include human participants, informed consent was not required.

Footnotes

Authorship Contributions: Concept - D.T.K., B.Y., A.Ö.A., O.G., M.D., F.F.K.; Design - D.T.K., B.Y., M.D., F.F.K.; Data Collection or Processing - D.T.K., B.Y., Ö.C.S., F.F.K.; Analysis or Interpretation - B.Y., Ö.C.S., O.G., M.D., F.F.K.; Literature Search - D.T.K., A.Ö.A., Ö.C.S.; Writing - D.T.K., Ö.C.S., F.F.K.

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. Abhisheka B, Biswas SK, Purkayastha B, Das D, Escargueil A. Recent trend in medical imaging modalities and their applications in disease diagnosis: a review. Multimed Tools Appl. 2024; 83: 43035-70.
2. Zangos S, Marquart F. Female reproductive system. In: Vogl TJ, Reith W, Rummeny EJ, editors. Diagnostic and interventional radiology. Berlin, Heidelberg: Springer; 2016. p. 929-55.
3. Iraha Y, Okada M, Iraha R, Azama K, Yamashiro T, Tsubakimoto M, et al. CT and MR imaging of gynecologic emergencies. Radiographics. 2017; 37: 1569-86.
4. Miao KH, Miao JH, Rosberger S, Dachman AH, Taouli B, Lewis SC. Advances in imaging and diagnosis of emphysematous cholecystitis. Healthcare (Basel). 2026; 14: 617.
5. Raman SP, Mahesh M, Blasko RV, Fishman EK. CT scan parameters and radiation dose: practical advice for radiologists. J Am Coll Radiol. 2013; 10: 840-6.
6. Skrzypek M, Wdowiak A, Panasiuk L, Stec M, Szczygieł K, Zybała M, et al. Effect of ionizing radiation on the female reproductive system. Ann Agric Environ Med. 2019; 26: 606-16.
7. Ray K. Toxicity of radiation: biological effects of ionizing radiation exposure on reproduction. In: Gupta RC, editor. Reproductive and developmental toxicology. 2nd ed. Cambridge (MA): Academic Press; 2017. p. 359-75.
8. Ray K, Choudhuri R. Effects of radiation on the reproductive system. In: Gupta RC, editor. Reproductive and developmental toxicology. Cambridge (MA): Academic Press; 2011. p. 291-9.
9. Mohammed Amin SS, Faraj KA. Review: influence of radiation on female fertility and pregnancy. Kirkuk Univ J Sci Stud. 2021; 16: 1-23.
10. Peiro A, Chegeni N, Danyaei A, Fatahiasl J, Tahmasbi M. Pelvic radiation dose measurement for trauma patients in multifield radiographic examinations: a phantom-based TLD dosimetry study. Health Sci Rep. 2023; 6: e1424.
11. Chu KH, Lin YT, Hsu CC, Chen CY, Pan LK. Evaluation of effective dose for a patient under Ga-67 nuclear examination using TLD, water phantom and a simplified model. J Radiat Res. 2012; 53: 989-98.
12. Senthilkumar S. Design of homogeneous and heterogeneous human equivalent thorax phantom for tissue inhomogeneity dose correction using TLD and TPS measurements. Int J Radiat Res. 2014; 12: 179-88.
13. Schaly B, Varchena V, Au P, Pang G. Evaluation of an anthropomorphic male pelvic phantom for image-guided radiotherapy. Rep Med Imaging. 2009; 2: 69-78.
14. Puntambekar S, Manchanda R. Surgical pelvic anatomy in gynecologic oncology. Int J Gynaecol Obstet. 2018; 143 Suppl 2 :86-92.
15. Standring S, editor. Gray's anatomy: the anatomical basis of clinical practice. 41st ed. Amsterdam: Elsevier; 2016.
16. Dalley AF, Agur AMR. Moore's Clinically Oriented Anatomy. 9th ed. Philadelphia: Wolters Kluwer Health; 2023.
17. Saksouk FA, Johnson SC. Recognition of the ovaries and ovarian origin of pelvic masses with CT. Radiographics. 2004; 24 Suppl 1: S133-46.
18. Mahadevan V. Anatomy of the lower urinary tract. Surgery (Oxford). 2019; 37: 351-8.
19. Obed RI, Ogbole GI, Majolagbe BS. Radiation doses to the uterus and ovaries in abdominopelvic computed tomography in a Nigerian tertiary hospital. West Afr J Radiol. 2025; 23: 7-11.
20. Gao Y, Mahmood U, Liu T, Quinn B, Gollub MJ, Xu XG, et al. Patient-specific organ and effective dose estimates in adult oncologic CT. AJR Am J Roentgenol. 2020; 214: 738-46.
21. Shubayr N, Alashban Y. Estimation of radiation doses and lifetime attributable risk of radiation-induced cancer in the uterus and prostate from abdomen pelvis CT examinations. Front Public Health. 2023; 10: 1094328.

İstanbul Med J 2026; 27(2): 122-8
DOI: 10.4274/imj.galenos.2026.48347

Is a Calcar-Replacement Femoral Stem as Effective as a Diaphyseal-Fixation Modular Stem in Hemiarthroplasty for Unstable Intertrochanteric Fractures?

Bekir Karagöz[1], Mustafa Erdem[2], Ali Geçer[3], Hünkar Çağdaş Bayrak[4], Alican Barış[5], İsmail Ağır[6]

[1]University of Health Sciences Türkiye, Eskişehir City Hospital, Clinic of Orthopedics and Traumatology, Eskişehir, Türkiye
[2]Afyonkarahisar State Hospital, Clinic of Orthopedics and Traumatology, Afyon, Türkiye
[3]University of Health Sciences Türkiye, Haydarpaşa Numune Training and Research Hospital, Clinic of Orthopedics and Traumatology, İstanbul, Türkiye
[4]Çekirge State Hospital, Clinic of Orthopedics and Traumatology, Bursa, Türkiye
[5]University of Health Sciences Türkiye, Physical Therapy and Rehabilitation Training and Research Hospital, Clinic of Orthopedics and Traumatology, İstanbul, Türkiye
[6]Adıyaman Training And Research Hospital, Clinic of Orthopedics and Traumatology, Adıyaman, Türkiye

ABSTRACT

Introduction: This study aims to compare the clinical and radiographic outcomes of cementless calcar replacement stems (CRSs) and distally fixed modular stems (DFMSs) in elderly patients undergoing hemiarthroplasty for unstable intertrochanteric femur fractures.

Methods: A retrospective cohort study was conducted on 138 patients who underwent cementless hip hemiarthroplasty for unstable intertrochanteric fractures at two tertiary university hospitals between 2017 and 2023. Patients aged 65 years or older with Arbeitsgemeinschaft für Osteosynthesefragen/Orthopaedic Trauma Association type A2.2, A2.3, or A3 fractures were included. Based on the type of implant, patients were divided into two groups: CRS (n=67) and DFMS (n=71). The groups were compared in terms of demographic data, operative time, intraoperative blood loss, transfusion requirements, complications, Harris Hip Score (HHS), and Parker Mobility Score (PMS). Minimal Clinically Important Difference (MCID) analyses were performed for functional outcomes.

Results: Hospital stay, operative time, intraoperative bleeding, and transfusion needs were significantly lower in the DFMS group than in the CRS group (p=0.002; p=0.004; p=0.024; p=0.003, respectively). The incidence of dislocation was higher in the CRS group (p=0.039), while no significant differences were observed in other complications. HHS and PMS scores at 6 and 12 months postoperatively did not differ significantly between groups, and none of the differences exceeded the MCID thresholds.

Conclusion: Although functional outcomes were comparable between the two stem types, DFMSs demonstrated statistically significant advantages in terms of surgical duration, blood loss, transfusion requirements, and complication rates.

Keywords: Intertrochanteric fracture, hemiarthroplasty, calcar replacement stem, modular stem, elderly patient, hip fracture

Introduction

Unstable intertrochanteric femur fractures are common in elderly individuals and are associated with significant morbidity and mortality (1). In such patients, the primary treatment goal is to minimize surgical complications and enable early mobilization (2,3). Currently, two main surgical options are employed: osteosynthesis and hip arthroplasty (4). Although osteosynthesis is generally effective in stable fractures, its limitations in elderly and osteoporotic patients, such as implant failure and restricted weight-bearing capacity, have led to increasing interest in arthroplasty as an alternative treatment option, particularly in unstable fracture patterns (5).

Femoral stems used in hip arthroplasty are available in various designs, including proximal (metaphyseal) or distal (diaphyseal) fixation, short or long length, and cemented or cementless fixation (2,6,7). While proximally fixed stems are typically favored in primary cases, there has been a growing trend toward the use of distally fixed modular stems (DFMSs) in elderly patients with unstable intertrochanteric fractures (8,9). These modular stems offer increased intraoperative flexibility, particularly in adjusting limb length and femoral anteversion (9,10). However, the optimal implant choice in cases of calcar deficiency remains controversial. Reconstruction of the calcar in such fractures is technically demanding and has been linked to increased complication rates, thus it is recommended by only a few authors (10). Some studies report that

Address for Correspondence: Bekir Karagöz, University of Health Sciences Türkiye, Eskişehir City Hospital, Clinic of Orthopedics and Traumatology, Eskişehir, Türkiye
E-mail: drbkr71@gmail.com **ORCID ID:** orcid.org/0000-0002-7447-452X

Cite this article as: Karagöz B, Erdem M, Geçer A, Bayrak HÇ, Barış A, Ağır İ. Is a calcar-replacement femoral stem as effective as a diaphyseal-fixation modular stem in hemiarthroplasty for unstable intertrochanteric fractures? İstanbul Med J. 2026; 27(2): 122-8

Received: 13.02.2026
Accepted: 30.03.2026
Publication Date: 12.05.2026

calcar replacement stems (CRSs) provide effective stabilization, while others suggest that d DFMSs are associated with lower complication rates (9-12).

The present study compares the clinical and radiographic outcomes of the cementless CRS and the DFMS in patients undergoing hemiarthroplasty for unstable intertrochanteric fractures. We hypothesize that the type of femoral stem used may have a significant impact on clinical, surgical, and radiographic outcomes in this population.

Methods

Study Design and Participants

This retrospective cohort study was approved by the Non-Interventional Clinical Research Ethics Committee of Adıyaman University (approval number: 2022/5-8, date: 24.05.2022) and conducted in accordance with the Declaration of Helsinki. The requirement for informed consent was waived by the ethics committee due to the retrospective design of the study. Patients who underwent cementless hip hemiarthroplasty for unstable intertrochanteric femur fractures between January 2017 and November 2023 at two tertiary referral centers (Adıyaman University Training and Research Hospital and Univeristy of Health Sciences Türkiye, Haydarpaşa Numune Training and Research Hospital) were included. Inclusion criteria were: age ≥65 years, diagnosis of an unstable intertrochanteric femur fracture classified as type A2.2, A2.3, or A3 according to the Arbeitsgemeinschaft für Osteosynthesefragen/ Orthopaedic Trauma Association (AO/OTA) classification, and treatment with either a cementless CRS or a DFMS. Exclusion criteria included age <65 years (n=15), history of hip fracture (n=11), pathological fractures (n=5), fractures resulting from high-energy trauma (n=8), multiple trauma (n=4), and use of any implant other than a CRS or DFMS (n=31). After excluding 74 patients based on these criteria, 138 patients met the inclusion criteria and were enrolled in the study (Figure 1). Among the included patients, 65 (47.1%) were male and 73 (52.9%) were female, with a mean age of 81.49±7.53 years. Based on the type of femoral stem used, 67 patients were assigned to the CRS group and 71 to the DFMS group.

Surgical Technique

A single surgical team experienced in hip arthroplasty performed all surgical procedures andworked collaboratively at both centers for an extended period. All patients received low-molecular-weight heparin for antithrombotic prophylaxis and 2 grams of intravenous cefazolin approximately 30 minutes before surgery for antibiotic prophylaxis. Surgeries were carried out under either spinal or general anesthesia. In all cases, the hip joint was accessed using a posterolateral approach with the patient positioned in the lateral decubitus position (13). After removal of the femoral head and neck fragments, the femoral canal was sequentially broached and rasped to prepare for implant insertion. Patients in the CRS group underwent cementless bipolar hemiarthroplasty with a CRS (TST SAN, İstanbul, Türkiye), while those in the DFMS group underwent a cementless long-stem modular bipolar hemiarthroplasty (Tıpmed, İstanbul, Türkiye) (Figures 2 and 3). Implant selection was based on predefined intraoperative considerations, including the presence of a significant calcar defect, degree of medial cortical comminution, extension of the fracture into the subtrochanteric region, bone quality, and femoral canal morphology. CRSs were generally preferred in cases with substantial calcar deficiency, severe medial cortical comminution, or inadequate proximal femoral support that could compromise metaphyseal fixation. In contrast, DFMSs were favored for cases with sufficient diaphyseal bone stock, particularly for fractures extending into the subtrochanteric region, where stable distal fixation could be achieved. Final implant positioning was confirmed intraoperatively via fluoroscopy. After hip reduction, joint stability and soft tissue tension were assessed manually, and adjustments to stem anteversion and leg length were made if necessary. A hemovac drain was placed in the surgical site, and the joint capsule was primarily closed. In cases of a detached greater trochanter fragment, the fragment was reduced and fixed with non-absorbable sutures or cerclage wire.

Postoperative Management and Rehabilitation

All patients followed a standardized postoperative rehabilitation protocol, harmonized across both centers and including both passive and active exercises. Hemovac drains placed at the surgical site were removed on the first postoperative day prior to mobilization. On the same day, patients began quadriceps-strengthening exercises and were allowed to bear partial weight on the operated limb with the aid of a walker, as tolerated. Intravenous cefazolin was administered for 24 hours postoperatively. For thromboembolism prophylaxis, low-molecular-weight heparin was continued until the sixth postoperative week. Anteroposterior radiographs of the hip were obtained shortly after surgery to assess early implant positioning. Patients without early complications were discharged between postoperative days 4 and 6, depending on their mobility. Routine clinical and radiographic follow-ups were scheduled at postoperative week 2 and at months 1, 3, 6, and 12, followed by annual outpatient visits.

Data Collection and Assessment Tools

Demographic and clinical data for all included patients were reviewed retrospectively from the hospital's electronic medical records. Collected variables included age, sex, fracture side, length of hospital stay, comorbidities, American Society of Anesthesiologists (ASA) score, type of anesthesia, and need for postoperative intensive care unit admission (14). Fracture types were classified preoperatively using the AO/OTA system (15). Classifications were independently performed by two orthopedic surgeons who were blinded to treatment allocation. In case of disagreement, the final classification was determined by consensus. Surgical data included operative time, intraoperative blood loss, and the number of transfused red blood cell units. Operative time was defined as the duration from skin incision to final skin closure. Blood loss was calculated as the sum of the volume in the suction container and the difference in weight between dry and used surgical gauze. The same type of gauze and the same measurement method were used across all cases. The transfusion requirement was recorded as the total number of red blood cell units administered to each patient. Functional outcomes were assessed using the Harris Hip Score (HHS) and the Parker Mobility Score (PMS) (16,17). Both scores were recorded preoperatively and at 6 and 12 months postoperatively. Preoperative scores were recorded based on the patients' premorbid functional status, as documented in medical records or obtained from patient and/or caregiver reports

at admission. HHS is a 100-point scale that evaluates pain, function, deformity, and range of motion; higher scores indicate better hip function (16). PMS evaluates a patient's mobility level and has been validated as a predictor of postoperative mortality (17). All perioperative surgical and medical complications were recorded. These included prosthesis dislocation, periprosthetic fracture, infection, deep vein thrombosis (DVT), osteolysis, and heterotopic ossification. Osteolysis was assessed on standard anteroposterior hip radiographs taken at or after the 12-month follow-up using the Gruen zone method (18). Heterotopic ossification was graded according to the Brooker classification (19). All HHS and PMS evaluations, as well as osteolysis and Brooker assessments, were performed by two independent orthopedic surgeons who were blinded to implant type. In cases of disagreement, a third reviewer was consulted to reach consensus.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, Version 25.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics were presented as means, standard deviations, frequencies, and percentages. The distribution of continuous variables was assessed using the Shapiro–Wilk test, which indicated that the data were not normally distributed. Between-group comparisons of independent continuous variables were conducted using the Mann–Whitney U test, while paired data were analyzed with the Wilcoxon signed-rank test. Categorical variables were compared using Fisher's exact test for variables with two levels and the chi-square test for variables with more than two levels. A p value <0.05 was considered statistically significant for all tests. To assess the effect size of statistically significant findings, Cohen's d was calculated. The mean effect size was found to be 0.93. Based on this value, a post-hoc

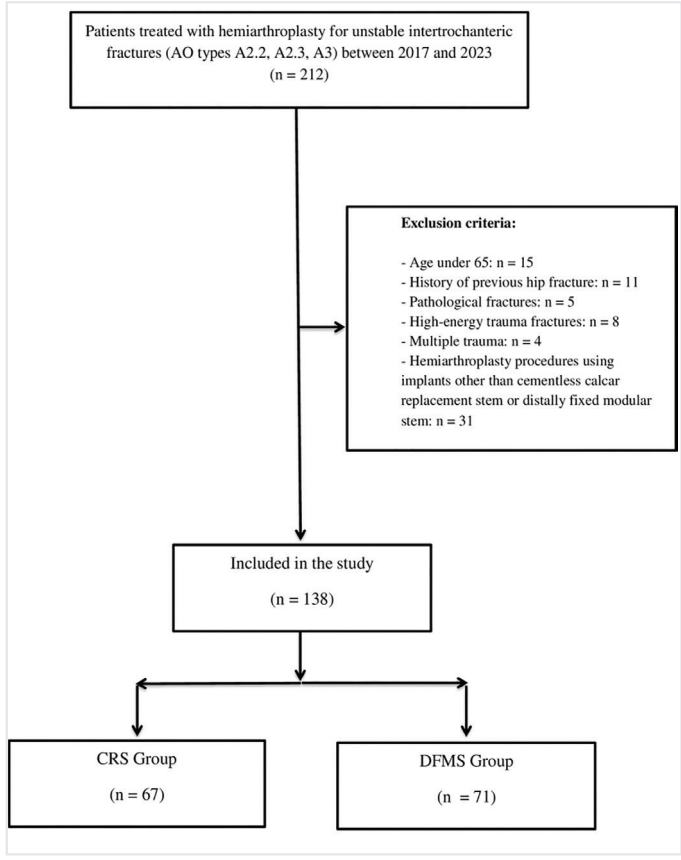

Figure 1. Flowchart of patient selection. AO: Arbeitsgemeinschaft für Osteosynthesefragen, CRS: Calcar replacement stem, DFMS: Distally fixed modular stem

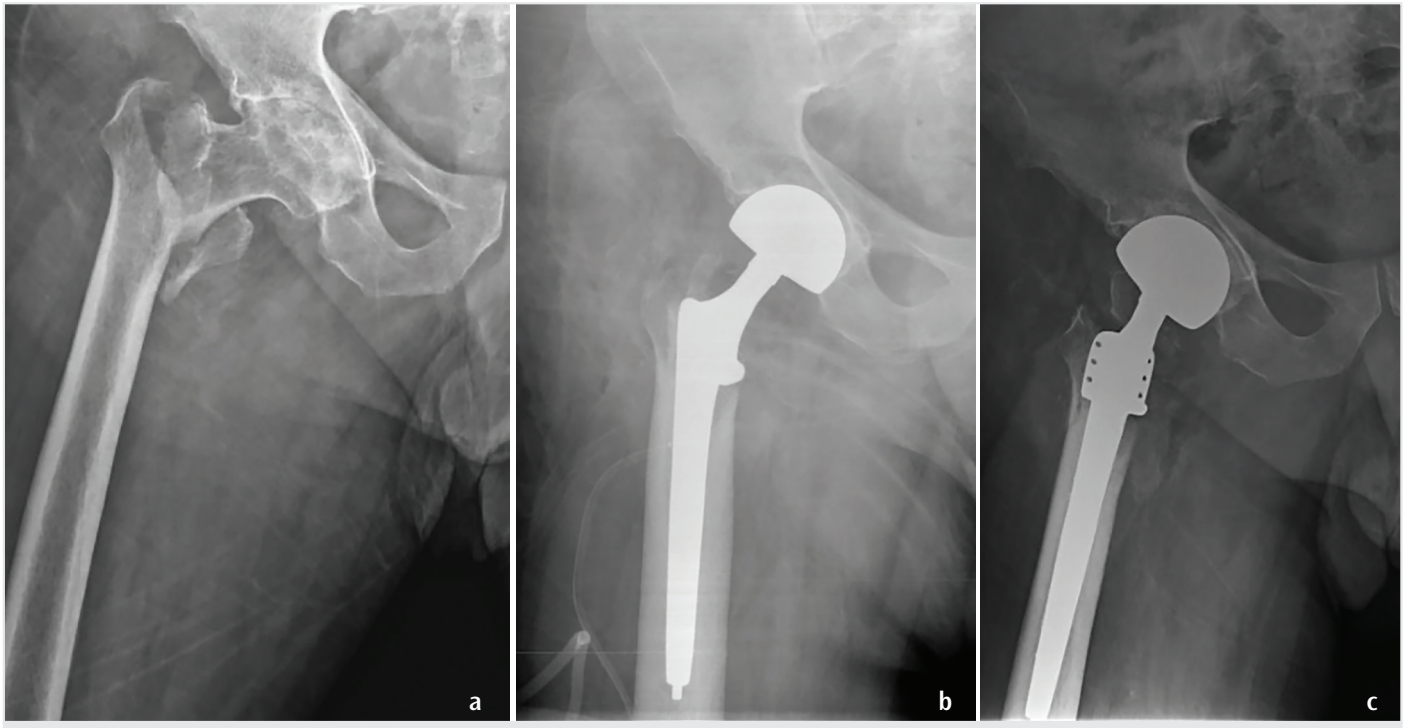

Figure 2. A 78-year-old male patient with an unstable intertrochanteric fracture of the right hip. Preoperative anteroposterior radiograph a) immediate postoperative radiograph following cementless hemiarthroplasty using a calcar replacement stem b), and radiograph at 1-year postoperative follow-up c). Consent of the patient is taken

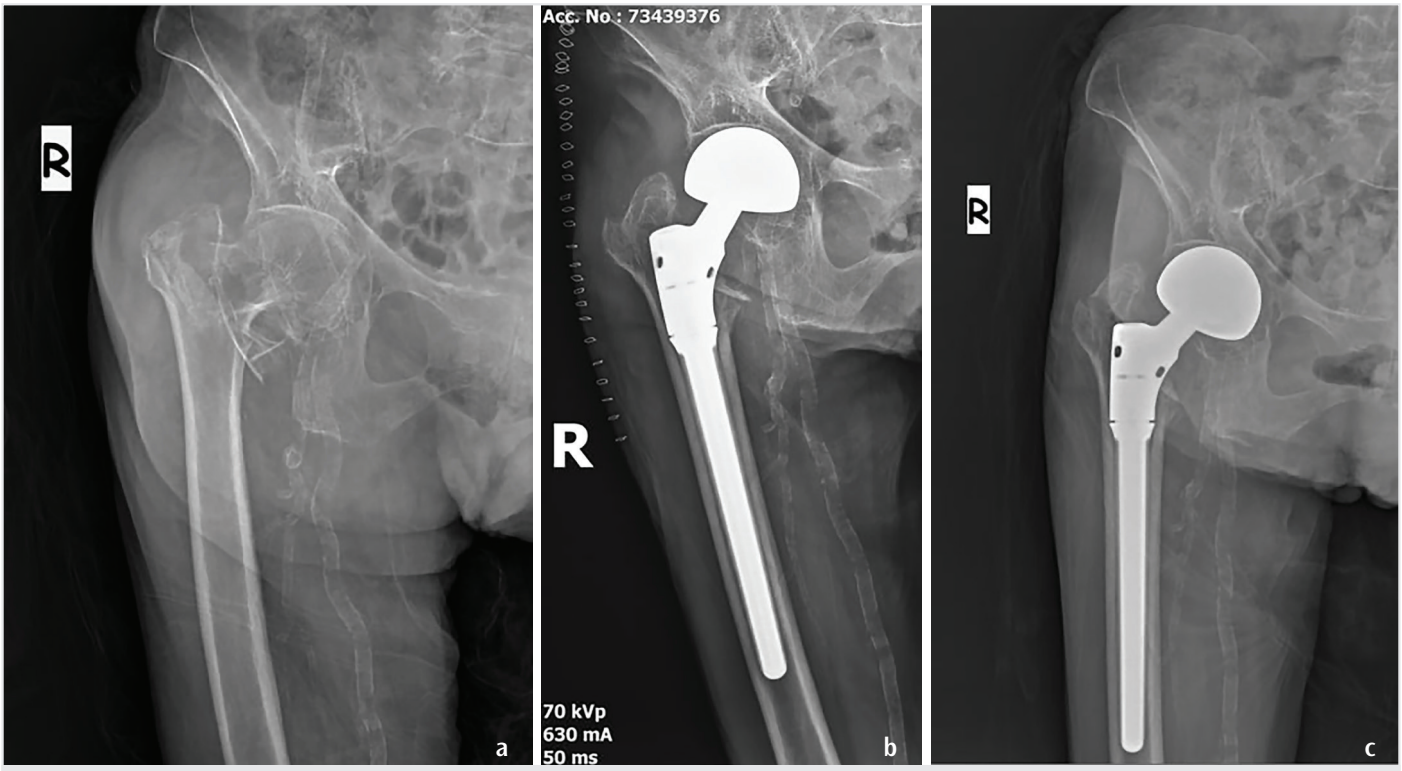

Figure 3. An 82-year-old female patient with an unstable intertrochanteric fracture of the right hip. Preoperative anteroposterior radiograph a), immediate postoperative radiograph following cementless hemiarthroplasty using a distally fixed modular stem b), and radiograph at 1-year postoperative follow-up). Consent of the patient is taken

power analysis indicated that the statistical power of the study exceeded 99% for the given sample size. To further evaluate changes in functional outcomes, minimal clinically important difference (MCID) values were calculated for the HHS and PMS. The MCID was determined using a distribution-based approach, defined as 50% of the pooled standard deviation of both groups (20).

Results

The demographic characteristics of the 138 patients included in the study are summarized in Table 1. Except for the length of hospital stay, no statistically significant differences were observed between the groups in terms of age, sex, side of fracture, AO/OTA classification, comorbidities, ASA score, type of anesthesia, or intensive care unit admission. The length of hospital stay was significantly shorter in the DFMS group than in the CRS group (p=0.002).

Significant differences were observed between the groups in mean operative time, intraoperative blood loss, and the number of transfused blood units (Table 2). The average operative time was significantly shorter in the DFMS group than in the CRS group (p=0.004).

Both intraoperative blood loss and transfusion volume were lower in the DFMS group than in the CRS group (p=0.001 and p=0.003, respectively). The proportion of patients requiring three or more units of blood transfusion was significantly higher in the CRS group (p=0.013).

Among all postoperative complications, the most frequently observed complication was prosthesis dislocation, with an overall incidence of 7.2%. This complication occurred significantly more often in the CRS group (p=0.039). No statistically significant differences were found between the groups for other complications (Table 2).

When functional outcomes were evaluated, no statistically significant differences were found between the CRS and DFMS groups in HHS or PMS at the preoperative, 6-month, or 12-month time points (Table 3). However, in both groups, postoperative HHS values were noticeably lower than preoperative scores. Despite these changes over time, the differences in mean scores between preoperative and postoperative assessments did not exceed the predefined MCID thresholds for either HHS or PMS and were therefore considered clinically insignificant.

Discussion

This study was conducted to compare the clinical and surgical outcomes of cementless hemiarthroplasty using CRS and DFMS in elderly patients with unstable intertrochanteric femoral fractures. The analysis showed that the length of hospital stay was significantly shorter in the DFMS group. Similarly, operative time, intraoperative blood loss, and transfusion requirements were significantly lower in the DFMS group. On the other hand, the incidence of prosthesis dislocation was significantly higher in the CRS group. However, no statistically significant differences were observed between the two groups with respect to functional outcomes as assessed by HHS and PMS scores at 6 and 12 months postoperatively.

One of the main challenges in the surgical treatment of unstable intertrochanteric fractures is managing defects in the femoral calcar region. In hemiarthroplasty, reconstruction of this area is important to prevent stem subsidence and reduce limb length discrepancies (7,11,21-

Table 1. Baseline demographic and clinical characteristics of the study groups

Variable	CRS Group (n=67)	DFMS Group (n=71)	p value
Age, mean ± SD	81.10±7.83	81.84±7.27	0.664[1]
Gender, n (%)			
Female	35 (52.2%)	38 (53.5%)	0.508[2]
Male	32 (47.8%)	33 (46.5%)	
Side, n (%)			
Right	38 (56.7%)	38 (53.5%)	0.419[2]
Left	29 (43.3%)	33 (46.5%)	
AO classification, n (%)			
A22	31 (46.3%)	30 (42.3%)	0.893[3]
A23	21 (31.3%)	18 (25.4%)	
A33	15 (22.4%)	17 (23.9%)	
Length of hospital stay, days	8.99 ± 2.86	6.46 ± 2.10	0.002[1]
Comorbidities, n (%)			
Hypertension	44 (65.7%)	44 (62.0%)	0.392[2]
Diabetes mellitus	21 (31.3%)	22 (31.0%)	0.555[2]
Coronary artery disease	7 (10.4%)	11 (15.5%)	0.266[2]
Alzheimer's disease	12 (17.9%)	13 (18.3%)	0.542[2]
COPD	8 (11.9%)	9 (12.7%)	0.454[2]
Chronic kidney disease	8 (11.9%)	9 (12.7%)	0.905[2]
ASA score, n (%)			
2	7 (10.4%)	10 (14.1%)	0.799[3]
3	46 (68.7%)	48 (67.6%)	
4	14 (20.9%)	13 (18.3%)	
Type of anesthesia, n (%)			
General	14 (20.9%)	16 (22.5%)	0.515[2]
Spinal	53 (79.1%)	55 (77.5%)	
Postoperative intensive care unit stay, n (%)	13 (19.4%)	10 (14.1%)	0.206[2]

[1]Mann–Whitney U test, [2]Fisher's exact test, [3]chi-square test, AO: Arbeitsgemeinschaft für Osteosynthesefragen, COPD: Chronic obstructive pulmonary disease, ASA: American Society of Anesthesiologists physical status classification, CRS: Calcar replacement stem, DFMS: Distally fixed modular stem, SD: Standard deviation

Table 2. Surgical outcomes and complication rates in CRS and DFMS groups

Variable	CRS group (n=67)	DFMS group (n=71)	p value
Operative time (minute)	164.85±29.21	123.87±12.05	0.004[1]
Intraoperative blood loss (mL)	914.62±160.20	841.26±93.37	0.001[1]
Blood transfusion (units)	2.33±0.88	1.90±0.74	0.003[1]
Blood transfusion, n (%)			0.013[3]
1 unit	12 (17.9%)	21 (29.6%)	
2 units	27 (40.3%)	38 (53.5%)	
3 units	22 (32.8%)	10 (14.1%)	
4 units	6 (9.0%)	2 (2.8%)	
Complications, n (%)			
Dislocation	8 (11.9%)	2 (2.8%)	0.039[2]
Periprosthetic fracture	5 (7.5%)	2 (2.8%)	0.197[2]
DVT	3 (4.5%)	2 (2.8%)	0.472[2]
Infection	4 (6.0%)	3 (4.2%)	0.468[2]
Osteolysis	3 (4.5%)	2 (2.8%)	0.472[2]
Heterotopic ossification	1 (1.5%)	0	0.486[2]

[1]Mann–Whitney U test, [2]Fisher's exact test, [3]chi-square test, DVT: Deep vein thrombosis, CRS: Calcar replacement stem; DFMS: Distally fixed modular stem

Table 3. Comparison of functional outcomes between patients treated with CRS and DFMS

Variable	CRS group (n=67)	DFMS group (n=71)	p value
Harris Hip Score			
Preoperative	77.55±6.18	76.25±6.76	0.379
Postoperative 6th month	56.09±11.37	57.85±5.77	0.102
Postoperative 12th month	66.69±8.58	65.87±8.15	0.350
Parker Mobility Score			
Preoperative	7.12±0.86	6.97±1.09	0.186
Postoperative 6th month	4.60±1.06	4.80±1.15	0.432
Postoperative 12th month	6.01±1.83	6.24±1.01	0.449

CRS: Calcar replacement stem, DFMS: Distally fixed modular stem

23). However, calcar reconstruction often requires additional soft tissue dissection and supplementary implants, which may increase operative time and intraoperative blood loss (9,10). For this reason, CRSs have been recommended to provide medial support when reconstruction is not feasible (7,11,24). Nevertheless, achieving stable fixation with these stems can be challenging due to poor bone quality, trochanteric fragmentation, and insufficient stabilization (2), and the technique may involve a considerable learning curve with limited implant availability in some centers (6,11). As an alternative, DFMSs have gained attention in cases of insufficient calcar support. These implants achieve fixation in the distal femoral diaphysis, where cortical bone is typically stronger, and their modular design allows for more precise adjustment of femoral anteversion and limb length (8,10,25,26). These features may facilitate intraoperative management and improve technical control. In our study, DFMS demonstrated significant advantages over CRS in terms of operative time, intraoperative blood loss, and transfusion requirements.

All of these parameters were statistically significant, suggesting a potential clinical advantage. A shorter hospital stay in the DFMS group may be associated with several factors. Reduced operative time and lower intraoperative blood loss in this group may have contributed to faster postoperative recovery and earlier mobilization. In addition, more stable distal fixation may facilitate improved early weight-bearing capacity, which can further support earlier discharge. However, these associations should be interpreted cautiously, and further prospective studies are needed to better clarify the underlying mechanisms.

The type of complications that may occur after hemiarthroplasty can vary depending on the design of the implant and the surgical technique used (6,24).While prosthesis dislocation and surgical site infection are among the most frequently encountered complications, others such as femoral shaft fracture, leg length discrepancy, nonunion of the greater trochanter, stem subsidence, osteolysis, and heterotopic ossification are observed less commonly (2,9). Several studies have explored the

relationship between stem design and postoperative complications. For example, Tsai et al. (27) reported lower complication rates with distally fixed stems, while Karaali and Çiloğlu. (2) found that proximally fixed short stems were associated with higher rates of dislocation and osteolysis compared to distally fixed long stems. However, this association has not been consistently confirmed across all studies, and some reports have shown similar outcomes with non-modular stems (28). In our study, postoperative complications following hemiarthroplasty were also analyzed in detail. Prosthesis dislocation emerged as the most common complication. Although dislocation was the most frequent complication in both groups, it occurred less frequently in the DFMS group. No significant differences were observed between the groups regarding other complications. The modular structure of DFMS, which allows more precise adjustment of anteversion during surgery, may have contributed to prosthesis stability. This technical advantage might partially explain the lower dislocation rate observed in the DFMS group.

Study Limitations

This study has several important limitations. First, due to its retrospective design, patient groups were not randomized, and implant selection was based on surgeon preference and clinical judgment rather than on a standardized protocol. Although predefined intraoperative considerations were used to guide implant selection, the final decision depended on intraoperative assessment and the surgeon's judgment. Therefore, the potential for selection bias cannot be excluded. This may have led to imperfect group homogeneity and the potential for selection bias. Although the follow-up period was acceptable compared with existing literature, the feasibility of long-term follow-up in elderly populations is inherently limited and may therefore restrict its overall impact on clinical outcomes. Finally, although the sample size was small, it was comparable to those of similar studies in the literature. Nevertheless, stronger evidence is needed from future studies with larger sample sizes, prospective designs, longer follow-up durations, and standardized surgical protocols to confirm and expand upon these findings.

Conclusion

This study provided a comparative evaluation of two hemiarthroplasty techniques for the treatment of unstable intertrochanteric fractures in elderly patients. The findings demonstrated that, while both techniques yielded comparable functional outcomes, DFMS offered statistically significant advantages in surgical parameters, such as operative time, intraoperative blood loss, transfusion requirements, and complication rates. These findings suggest potential advantages of modular stem designs in this patient population.

Ethics

Ethics Committee Approval: The approval was obtained from the Non-Interventional Clinical Research Ethics Committee of Adıyaman University (approval number: 2022/5-8, date: 24.05.2022).

Informed Consent: The requirement for informed consent was waived by the ethics committee due to the retrospective design of the study.

Footnotes

Authorship Contributions:

Surgical and Medical Practices - B.K., A.G.; Concept - B.K., M.E., H.Ç.B., A.B., İ.A.; Design - B.K., M.E., A.G., A.B., İ.A.; Data Collection or Processing - M.E., A.G., H.Ç.B.; Analysis or Interpretation - B.K., A.G.; Literature Search - B.K., M.E., H.Ç.B.; Writing - B.K., H.Ç.B., A.B., İ.A.

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. Tseng FJ, Chia WT, Pan RY, Lin LC, Shen HC, et al. Comparison of arthroplasty vs. osteosynthesis for displaced femoral neck fractures: a meta-analysis. J Orthop Surg Res. 2017; 12: 131.
2. Karaali E, Çiloğlu O. Metaphyseal vs. diaphyseal fixed-stem hemiarthroplasty in treating unstable intertrochanteric fractures in elderly patients. Ulus Travma Acil Cerrahi Derg. 2021; 27: 104-8.
3. Luo X, He S, Zeng D, Lin L, Li Q. Proximal femoral nail antirotation versus hemiarthroplasty in the treatment of senile intertrochanteric fractures: case report. Int J Surg Case Rep. 2017; 38: 37-42.
4. Yu X, Wang H, Duan X, Liu M, Xiang Z. Intramedullary versus extramedullary internal fixation for unstable intertrochanteric fracture, a meta-analysis. Acta Orthop Traumatol Turc. 2018; 52: 299-307.
5. Socci AR, Casemyr NE, Leslie MP, Baumgaertner MR. Implant options for the treatment of intertrochanteric fractures of the hip: rationale, evidence, and recommendations. Bone Joint J. 2017; 99-B(1): 128-33.
6. Göçer H, Coşkun S, Karaismailoğlu N. Comparison of treatment of unstable intertrochanteric fracture with different arthroplasty methods. Niger Med J. 2016; 57: 81-5.
7. Cobden A, Camurcu Y, Duman S, Kocabiyik A, Kıs M, Saklavcı N. Mid-term survivals of cemented calcar-replacement bipolar hemiarthroplasty for unstable intertrochanteric fractures in elderly patients. Injury. 2019; 50: 2277-81.
8. Cohn MR, Tetreault MW, Li J, Kunze KN, Nahhas CR, Michalski JF, et al. Is there a benefit to modularity for femoral revisions when using a splined, tapered titanium stem? J Arthroplasty. 2020; 35: S278-83.
9. Çiloğlu O, Karaali E, Kuşvuran Özkan A, Ekiz T. Distally-fixed non-modular monoblock fluted long-stem hemiarthroplasty versus proximal femoral nailing for elderly patients with an osteoporotic intertrochanteric fracture: a retrospective comparative study. Hip Int. 2022; 32: 124-30.
10. Zha GC, Liu J, Wang Y, Feng S, Chen XY, Guo KJ, et al. Cementless distal fixation modular stem without reconstruction of femoral calcar for unstable intertrochanteric fracture in patients aged 75 years or more. Orthop Traumatol Surg Res. 2019; 105: 35-9.
11. Wada K, Mikami H, Oba K, Yonezu H, Sairyo K. Cementless calcar-replacement stem with integrated greater trochanter plate for unstable intertrochanteric fracture in very elderly patients. J Orthop Surg (Hong Kong). 2017; 25: 2309499016684749.
12. Khatri K, Banga RK, Malhotra N, Bansal D. Cemented calcar replacement versus long stem cemented hemiarthroplasty in unstable ıntertrochanteric fractures in octogenarians. Rev Bras Ortop (Sao Paulo). 2021; 57: 511-20.

13. Foran JRH, Della Valle CJ. Posterolateral approach to the hip. In: Nho SJ, Bedi A, Salata MJ, Mather RC III, Kelly BT, editors. Hip arthroscopy and hip joint preservation surgery. Cham: Springer; 2022.

14. Quach LH, Jayamaha S, Whitehouse SL, Crawford R, Pulle CR, Bell JJ. Comparison of the Charlson Comorbidity Index with the ASA score for predicting 12-month mortality in acute hip fracture. Injury. 2020; 51: 1004-10.

15. Meinberg EG, Agel J, Roberts CS, Karam MD, Kellam JF. Fracture and dislocation classification compendium-2018. J Orthop Trauma. 2018; 32 Suppl 1: S1-170.

16. Vishwanathan K, Akbari K, Patel AJ. Is the modified Harris hip score valid and responsive instrument for outcome assessment in the Indian population with pertrochanteric fractures? J Orthop. 2018; 15: 40-6.

17. Voeten SC, Nijmeijer WS, Vermeer M, Schipper IB, Hegeman JH; DHFA Taskforce study group. Validation of the Fracture Mobility Score against the Parker Mobility Score in hip fracture patients. Injury. 2020; 51: 395-9.

18. Molloy DO, Munir S, Jack CM, Cross MB, Walter WL, Walter WK Sr. Fretting and corrosion in modular-neck total hip arthroplasty femoral stems. J Bone Joint Surg Am. 2014; 96: 488-93.

19. Brooker AF, Bowerman JW, Robinson RA, Riley LH Jr. Ectopic ossification following total hip replacement. Incidence and a method of classification. J Bone Joint Surg Am. 1973; 55: 1629-32.

20. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008; 61: 102-9.

21. Liu P, Jin D, Zhang C, Gao Y. Revision surgery due to failed internal fixation of intertrochanteric femoral fracture: current state-of-the-art. BMC Musculoskelet Disord. 2020; 21: 573.

22. Bonnevialle P, Saragaglia D, Ehlinger M, Tonetti J, Maisse N, Adam P, et al. Trochanteric locking nail versus arthroplasty in unstable intertrochanteric fracture in patients aged over 75 years. Orthop Traumatol Surg Res. 2011; 97: S95-100.

23. Abdelkhalek M, Ali AM, Abdelwahab M. Cemented bipolar hemiarthroplasty with a cerclage cable technique for unstable intertrochanteric hip fractures in elderly patients. Eur J Orthop Surg Traumatol. 2013; 23: 443-8.

24. Chammout G, Kelly-Pettersson P, Hedbeck CJ, Stark A, Mukka S, Sköldenberg O. HOPE-trial: hemiarthroplasty compared with total hip arthroplasty for displaced femoral neck fractures in octogenarians: a randomized controlled trial. JB JS Open Access. 2019; 4: e0059.

25. Perticarini L, Rossi SMP, Fioruzzi A, Jannelli E, Mosconi M, Benazzo F. Modular tapered conical revision stem in hip revision surgery: mid- term results. BMC Musculoskelet Disord. 2021; 22: 29.

26. Wiik AV, Aqil A, Al-Obaidi B, Brevadt M, Cobb JP. The impact of reducing the femoral stem length in total hip arthroplasty during gait. Arch Orthop Trauma Surg. 2021; 141: 1993-2000.

27. Tsai SW, Chen CF, Wu PK, Huang CK, Chen WM, Chang MC. Does implant selection impact postoperative complications following hip arthroplasty for failed intertrochanteric fractures? A retrospective comparative study. Artif Organs. 2016; 40: 798-804.

28. Regis D, Sandri A, Bartolozzi P. Stem modularity alone is not effective in reducing dislocation rate in hip revision surgery. J Orthop Traumatol. 2009; 10: 167-71.

İstanbul Med J 2026; 27(2): 129-35
DOI: 10.4274/imj.galenos.2026.32698

Relationship Between Cognitive Flexibility, Pregnancy-Related Anxiety, and Attentional Functions in Pregnant Women

Aslı Deşer, Sibel Tunç Karaman, Okcan Basat

University of Health Sciences Türkiye, Gaziosmanpaşa Training and Research Hospital, Clinic of Family Medicine, İstanbul, Türkiye

ABSTRACT

Introduction: This study aimed to examine cognitive flexibility and pregnancy-related anxiety among pregnant women and to explore their relationships with attentional functions during pregnancy.

Methods: A cross-sectional study was conducted with 125 pregnant women attending the Family Medicine Outpatient Clinic of Gaziosmanpaşa Training and Research Hospital. Data were collected using a descriptive information form, the cognitive flexibility inventory (CFI), the pregnancy-related anxiety scale (PRAS), and verbal fluency tests (phonemic and semantic).

Results: Participants had a mean age of 28.5±6.01 years (range: 18-44). The mean total number of words produced in verbal fluency tests was 70.95±24.0. The mean PRAS score was 70.86±14.20, reflecting low-to-moderate anxiety levels. Their mean CFI score was 68.65±10.66, suggesting moderate-to-good cognitive flexibility. PRAS and CFI scores were negatively correlated (r: –0.238, p=0.008), while CFI scores were positively correlated with total word count (r: 0.299, p=0.001). Participants who reported persistent attention difficulties and forgetfulness had significantly higher PRAS scores than those without these complaints (p=0.010 for attention difficulties and p=0.011 for forgetfulness).

Conclusion: Pregnant women in this study exhibited low-to-moderate pregnancy-related anxiety and moderate-to-good cognitive flexibility. Higher cognitive flexibility was significantly associated with lower anxiety and better verbal fluency. Although no direct link was found between anxiety and verbal fluency, self-reported attentional difficulties were associated with higher anxiety and reduced cognitive flexibility. These findings suggest that cognitive flexibility may act as a protective factor during pregnancy and may represent a promising target for prenatal mental health interventions.

Keywords: Anxiety, attention, cognitive flexibility, pregnancy

Introduction

Pregnancy represents a unique and complex period in a woman's life, involving profound biological, physiological, psychological, and social changes (1). Alongside the physical adaptations needed to sustain fetal development, women often face psychosocial stressors such as limited socioeconomic resources, work–family responsibilities, and adjustments to new maternal roles (2). These demands, coupled with hormonal fluctuations, can heighten emotional vulnerability; recent studies estimate that 15–25% of pregnant women worldwide experience clinically significant psychological distress, with similar rates reported in Türkiye (3).

Pregnancy-related anxiety is conceptually distinct from generalized anxiety disorders, encompassing concerns about the baby's health, childbirth complications, maternal appearance, and postpartum responsibilities (4). It has been linked to adverse maternal and neonatal outcomes, including preterm birth, impaired maternal–infant bonding, and postpartum depression. Such anxiety may also affect cognitive domains such as attention, memory, and executive functioning, underscoring the importance of resilience factors that could mitigate its impact (5).

Cognitive flexibility—the ability to shift perspectives, adapt strategies, and generate alternative responses—plays a critical role in adapting to change (6). Lower cognitive flexibility has been consistently reported in individuals with anxiety disorders compared to healthy controls (7). In pregnancy, higher flexibility may help women manage rapid and multidimensional changes, potentially reducing the psychological burden of anxiety (8). However, to the best of our knowledge, no study in Türkiye has directly examined the interplay between cognitive flexibility, pregnancy-related anxiety, and attentional functioning.

Many pregnant women also report cognitive changes, often described as "baby brain" or "placental brain". These self-perceived issues—reduced attention, forgetfulness, slower memory retrieval—are supported by some neuropsychological findings showing poorer performance in attention, memory, and visuospatial tasks among pregnant women compared with non-pregnant controls (9). Although objective and subjective findings do not always align, recent reviews indicate that such complaints are

Address for Correspondence: Assoc. Prof. Sibel Tunç Karaman, MD, University of Health Sciences Türkiye, Gaziosmanpaşa Training and Research Hospital, Clinic of Family Medicine, İstanbul, Türkiye
E-mail: drsibeltunc@hotmail.com **ORCID ID:** orcid.org/0000-0003-1833-8758

Cite this article as: Deşer A, Tunç Karaman S, Basat O. Relationship between cognitive flexibility, pregnancy-related anxiety, and attentional functions in pregnant women. İstanbul Med J. 2026; 27(2): 129-35

Received: 04.12.2025
Accepted: 01.04.2026
Publication Date: 12.05.2026

common among pregnant women, with a substantial proportion reporting changes in memory and attention (10). It is plausible that diminished cognitive flexibility could exacerbate attentional difficulties, whereas greater flexibility may provide a protective advantage.

Despite evidence linking cognitive flexibility to better coping and lower childbirth fear, few studies have integrated these domains to examine whether flexibility can buffer the cognitive effects of pregnancy-related anxiety (8,11). Addressing this gap could inform interventions to enhance maternal cognitive and emotional well-being. Therefore, the present study aimed to investigate the relationship between cognitive flexibility, pregnancy-related anxiety, and attentional functions by combining self-reports with objective neuropsychological measures to better understand cognitive–emotional adaptation during pregnancy.

Methods

Study Design, Setting, and Participants

This cross-sectional study was conducted at the Family Medicine Outpatient Clinic of Gaziosmanpaşa Training and Research Hospital between June 8 and September 30, 2023. The study included 125 volunteer pregnant women who met the inclusion criteria.

Eligible participants were pregnant women aged ≥18 years who had sufficient Turkish literacy and comprehension to complete the study and provided informed consent. Exclusion criteria included diagnosed psychiatric disorders affecting cognition, cognitive impairment, inability to cooperate, or hearing and speech impairments.

Sample Size Estimation

The sample size was calculated using G*Power version 3.1.9.4, based on effect size estimates from prior research. The required sample size was determined to be 108 participants (α: 0.05, power: 0.95). A total of 125 participants were recruited, exceeding the required minimum sample size.

Data Collection

Data were collected using a descriptive information form, the cognitive flexibility inventory (CFI), the pregnancy-related anxiety scale (PRAS), and verbal fluency tests (phonemic and semantic).

Descriptive Information Form

This form, developed by the researchers, included questions on sociodemographic characteristics (age, education level, employment status, income level), general medical and obstetric history (e.g., presence of chronic disease, gestational week, number of pregnancies, high-risk pregnancy status), and self-reported cognitive flexibility and attention. High-risk pregnancy encompasses pre-existing medical conditions, previous and current pregnancy complications, advanced maternal age, and multiple pregnancy.

Pregnancy-Related Anxiety Scale

The PRAS was developed by Brunton et al. (12) in 2018 and adapted into Turkish by Kurt and Arslan (13) in 2021. The 31 items are grouped into nine subscales and are rated on a four-point Likert scale from 1 ("Never") to 4 ("Very often"). Nine items are reverse-scored. The total score ranges from 31 to 124, with higher scores indicating greater pregnancy-related anxiety. The Cronbach's alpha was 0.92 for the original version and 0.89 for the Turkish version (12,13).

Cognitive Flexibility Inventory

The CFI was developed by Dennis and Vander Wal (14) in 2010 to assess the ability to generate alternative, adaptive, and appropriate responses in challenging situations. The Turkish adaptation and validation were conducted by Gulum and Dag (15) in 2012. The 20-item scale includes two subscales: "Alternatives" and "Control". Six items are reverse-scored. Total scores range from 20 to 100, with higher scores reflecting greater cognitive flexibility. Cronbach's alpha values for the "Alternatives" subscale were 0.91, and for the "Control" subscale, 0.86 and 0.84 in initial and final measurements (14,15).

Verbal Fluency Tests

Verbal fluency tests assess phonemic (lexical) and semantic (categorical) fluency. They are widely used in clinical neurological and cognitive assessments, including patients with dementia, Parkinson's disease, psychosis, and depression, as well as healthy individuals. These tests are advantageous because they are easy to administer, brief, and reliant on verbal skills rather than literacy.

In the semantic fluency test, participants are asked to name as many items as possible from a specific category (e.g., animals, fruits, supermarket items) within a limited time. In the phonemic fluency test, participants produce as many different words as possible starting with a specific letter. In the Turkish version, the letters "K", "A", and "S" are commonly used because they occur with high frequency in the language. Scoring can be conducted separately for each letter/category or as a total score (16,17).

Ethical Approval

The study protocol was approved by the Gaziosmanpaşa Training and Research Hospital Clinical Research Ethics Committee (approval number: 76, date: June 7, 2023). All procedures were carried out in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics version 22. The normality of continuous variables was assessed using the Kolmogorov–Smirnov test. Descriptive statistics (means, standard deviations, frequencies) were used to summarize the data. Group comparisons for normally distributed variables were conducted using one-way analysis of variance (post-hoc Tukey HSD test), while the Kruskal–Wallis test (post-hoc Dunn's test) was applied for non-normally distributed variables. An independent-samples t-test was used for two-group comparisons of normally distributed variables, and the Mann–Whitney U test was used for non-normally distributed variables. Correlation analyses were performed using Pearson's correlation coefficient for normally distributed variables and Spearman's rho for non-normally distributed variables. Internal consistency of the scales was assessed using Cronbach's alpha coefficient. Statistical significance was set at $p<0.05$.

Results

This cross-sectional study included 125 pregnant women aged 18–44 years (mean: 28.55±6.01). Gestational age averaged 20.58±9.09 weeks (range: 5–39), and the number of pregnancies averaged 2.38±1.49 (range: 1–5). Table 1 presents participants' sociodemographic, obstetric, and cognitive-experiential characteristics. Overall, 20.0% (n=25) of participants had a pre-pregnancy chronic disease; 22.4% (n=28) were classified as high-risk; and 39.2% and 38.4% reported frequent or constant attention difficulties and forgetfulness, respectively.

Table 2 summarizes the descriptive statistics for PRAS, CFI, and verbal fluency scores. PRAS averaged 70.86±14.20, indicating low-to-moderate pregnancy-related anxiety, while CFI averaged 68.65±10.66, reflecting moderate-to-high cognitive flexibility. Subscale data and reliability coefficients for both scales are provided in the table.

In verbal fluency performance, total word count (phonemic + semantic) was 70.95±24.0, with phonemic fluency at 32.39±12.48 and semantic fluency at 38.56±12.64.

As shown in Table 3, Pearson correlation analysis revealed a significant negative association between PRAS and CFI scores (r: –0.238, p=0.008). CFI scores were positively correlated with verbal fluency word count (r: 0.299, p=0.001). No significant correlation was found between PRAS and word count (p>0.05).

Table 4 presents a comparison of PRAS, CFI, and verbal fluency scores across sociodemographic, obstetric, and attention-related variables. Education level was significantly associated with both CFI (p=0.020) and word count (p=0.001), while income level was associated with CFI (p=0.001) and verbal fluency (p=0.001). Self-reported cognitive complaints were strongly associated with outcomes: frequent attention difficulties were linked to higher PRAS scores (p=0.004); frequent forgetfulness was associated with higher PRAS (p=0.010) scores, lower CFI scores (p=0.010), and lower verbal fluency scores (p=0.001).

Table 5 presents correlation analyses among clinical variables. Gestational age was positively correlated with CFI (r: 0.220, p=0.014), and the number of pregnancies was negatively correlated with verbal fluency (r: 0.256, p=0.004). Maternal age showed no significant associations with PRAS, CFI, or verbal fluency scores.

Table 1. Sociodemographic, obstetric, and cognitive experience characteristics of the participants (n=125)

Variables		n	%
Sociodemographic characteristics			
Education level	Literate	7	5.6
	Primary school	30	24
	Middle school	35	28
	High school	30	24
	University	23	18.4
Marital status	Single	9	7.2
	Married	116	92.8
Income level	Low	45	36
	Moderate	51	40.8
	High	29	23.2
Obstetric characteristics			
Pre-pregnancy chronic disease	No	100	80
	Yes	25	20
High-risk pregnancy	No	97	77.6
	Yes	28	22.4
Attending regular antenatal follow-up	No	37	29.6
	Yes	88	70.4
Attention- and memory-related experiences during pregnancy			
Self-reported attention difficulties during pregnancy	Never	16	12.8
	Rarely	60	48
	Often	39	31.2
	Always	10	8
Forgetfulness during pregnancy	Never	13	10.4
	Rarely	64	51.2
	Often	41	32.8
	Always	7	5.6

Data are presented as n (number) and % (percentage)

Table 2. Descriptive statistics for PRAS, CFI, and verbal fluency scores

	Min-max	Mean ± SD	Cronbach's alpha
PRAS			
Total score	34–112	70.86±14.20	0.888
Subscale scores			
Concerns about childbirth	6–24	15.22±5.06	0.904
Body image concerns	4–16	8.21±3.45	0.889
Attitudes toward childbirth	3–12	6.92±2.34	0.779
Concerns about motherhood	3–12	5.93±2.34	0.719
Acceptance of pregnancy	3–11	4.81±2.16	0.810
Anxiety indicators	3–12	8.27±2.44	0.835
Attitudes toward healthcare professionals	3–10	5.50±1.87	0.779
Avoidance	3–12	7.20±3.39	0.938
Concerns about the baby	3–12	8.80±2.05	0.814
CFI			
Total score	39–95	68.65±10.66	0.867
Subscale scores			
Control	8–33	18.43±4.99	0.714
Alternative	24–65	50.22±8.57	0.929
Verbal fluency test			
Total word count	31–156	70.95±24.00	-
Verbal fluency (total)	11–78	32.39±12.48	-
Letter "K"	5–31	11.81±4.44	-
Letter "A"	3–26	10.62±4.55	-
Letter "S"	3–26	9.96±4.39	-
Category fluency (total)	20–78	38.56±12.64	-
Animals	0–31	13.98±5.06	-
Fruits	4–24	9.42±4.04	-
Market items	7–31	15.15±5.19	-

Data presented as minimum (min), maximum (max), mean and standard deviation (SD) values. PRAS: Pregnancy-related anxiety scale, CFI: Cognitive flexibility inventory

Table 3. Correlation between pregnancy-related anxiety scale, cognitive flexibility inventory, and verbal fluency test (total word count) scores

		PRAS total score	CFI total score	Total word count
PRAS total score	r	1		
	p			
CFI total score	r	-0.238	1	
	p	**0.008***		
Total word count	r	0.077	0.299	1
	p	0.392	**0.001***	

Pearson correlation analysis. *p<0.05. PRAS: Pregnancy-related anxiety scale, CFI: Cognitive flexibility inventory

Table 4. Comparison of scale scores according to sociodemographic, obstetric and attention-related characteristics

		PRAS total score	CFI total score	Total word count
		Mean ± SD	Mean ± SD	Mean ± SD
Education level	Literate	66.86±5.87	64.43±9.73	56±17.05
	Primary school	69.03±17.33	65.97±8.96	61.8±24.18
	Middle school	70±14.33	66.37±12.29	63±18.24
	High school	73.37±14.01	71.07±10.88	73.27±18.17
	University	72.52±11.57	73.74±7.76	96.52±21.73
Marital status	p	[1]0.664	**[1]0.020***	**[2]0.001***
	Single	79.44±18.74	64.89±12.11	67.22±15.87
	Married	70.2±13.67	68.94±10.54	71.24±24.55
Income level	p	[3]0.060	[3]0.274	[4]0.778
	Low	70.47±16.41	64.69±10.52	59.18±17.36
	Moderate	70.47±14.18	68.67±10.83	72.25±22.69
	High	72.17±10.43	74.76±7.45	86.93±25.87
	p	[1]0.854	**[1]0.001***	**[2]0.001***
Pre-pregnancy chronic disease	No	70.97±14.17	69.03±10.71	70.69±23.01
	Yes	70.44±14.63	67.12±10.5	72±28.12
	p	[3]0.868	[3]0.425	[4]0.892
High-risk pregnancy	No	71.22±14.77	68.65±11.34	70.02±25.95
	Yes	70.54±13.77	68.65±10.07	71.82±22.22
	p	[3]0.654	[3]0.171	[4]0.457
Attending regular antenatal follow-up	No	69.29±15.53	66.97±11.03	64.03±18.67
	Yes	72.47±12.64	70.35±10.07	77.98±26.78
	p	[3]0.891	**[3]0.015***	**[4]0.023***
Self-reported attention difficulties during pregnancy	Never	64.00±16.54	70±14.34	78.44±19.5
	Rarely	68.77±13.30	69.35±10.09	68.52±21.73
	Often	74.21±11.08	68.05±9.89	70.85±28.18
	Always	81.40±19.03	64.6±10.73	74±26.61
	p	**[1]0.004***	[1]0.560	[2]0.286
Forgetfulness during pregnancy	Never	68.62±16.19	64.38±14.72	77.92±23.94
	Rarely	67.05±12.98	69.28±10.44	68.03±22.81
	Often	74.59±12.81	69.98±8.86	71.66±24.09
	Always	88.14±13.15	63.00±12.30	80.57±33.49
	p	[3]0.912	**[3]0.010***	**[4]0.001***

[1]One-way ANOVA test, [2]Kruskal-Wallis test, [3]Student's t-test, [4]Mann-Whitney U test, *p<0.05, Data presented as mean and standard deviation (SD) values. ANOVA: Analysis of variance, PRAS: Pregnancy-related anxiety scale, CFI: Cognitive flexibility inventory

Table 5. Correlation of scale scores with age, gestational age, and number of pregnancies

		PRAS total score	CFI total score	Total word count
Age	r	-0.089	0.088	0.095
	p	0.326	0.329	0.294
Gestational age	r	-0.112	**0.220**	0.079
	p	0.212	**0.014***	0.382
Number of pregnancies	r	-0.118	-0.050	**-0.256**
	p	0.189	0.580	**0.004***

*Pearson correlation analysis; p<0.05. PRAS: Pregnancy-related anxiety scale, CFI: Cognitive flexibility inventory

Discussion

This study examined the interrelationships between pregnancy-related anxiety, cognitive flexibility, and attentional functions in a cohort of pregnant women. Participants reported low-to-moderate anxiety and moderate-to-good cognitive flexibility. Greater flexibility was consistently linked with lower anxiety and better verbal fluency, suggesting that it may serve as a psychological resource during pregnancy. Anxiety, however, showed no significant association with verbal fluency.

The influence of sociodemographic factors on both psychological and cognitive outcomes revealed nuanced patterns. Maternal age, often considered a determinant of psychological well-being and obstetric risk, showed no significant association with anxiety or cognitive performance in this cohort. This finding is consistent with several regional studies, although some reports link advanced maternal age with elevated anxiety (18-20). This discrepancy may reflect differences in coping strategies among older mothers.

Beyond maternal age, other sociodemographic variables such as education and income also revealed complex patterns. Educational attainment and income showed no consistent associations with pregnancy-related anxiety in this cohort. While some Turkish studies reported mixed findings (21,22). International reviews identified lower education as a risk factor for prenatal anxiety (23). In our sample, higher education was linked to better cognitive flexibility and verbal fluency, echoing evidence that educational experiences strengthen cognitive reserve and problem-solving capacity (24). Income patterns were similarly heterogeneous, with cultural and family support buffering financial stress in lower-income groups, and heightened expectations creating pressures in higher-income groups (25). Together, these findings underscore how socioeconomic context shapes maternal psychological experiences in complex ways.

Obstetric variables also yielded relevant associations. Gestational week was positively associated with cognitive flexibility in our cohort, suggesting that adaptability may increase as pregnancy advances, possibly reflecting psychological adjustment to evolving maternal roles (24). However, gestational week was not significantly related to anxiety, partially consistent with evidence of trimester-specific fluctuations and with a recent Turkish study reporting that pregnancy-specific stress increased with advancing gestational week (26,27). In contrast, parity was negatively associated with verbal fluency, indicating that women with more pregnancies tended to produce fewer words in fluency tasks. This finding diverges from some reports linking primigravida status to higher anxiety rather than cognitive performance (9,11). Multiparous women may experience cumulative fatigue and increased caregiving responsibilities, which could contribute to reduced performance in verbal fluency tasks. These observations highlight the need for future studies in larger, diverse cohorts.

A particularly notable finding was the role of subjective forgetfulness, which was significantly associated with higher anxiety, reduced cognitive flexibility, and poorer verbal fluency. This highlights the close relationship between emotional and cognitive changes during pregnancy. Recent studies report modest declines in memory and attention during pregnancy, consistent with women's subjective complaints (10). While antenatal depression has been linked to attentional biases toward emotional stimuli, other work has found no objective impairments, suggesting a possible gap between perceived and actual performance (28). Such complaints may also reflect unmeasured factors, highlighting the need to combine subjective and objective assessments.

Overall, the results support cognitive flexibility as a potential psychological buffer. Higher cognitive flexibility was associated with reduced anxiety and better attentional performance, supporting theoretical models that view adaptable thinking as a resource for coping with stress (29). Mindfulness-based interventions, including digital applications, have been shown in meta-analyses to reduce perinatal depression and anxiety (30). Their integration into routine antenatal care may represent a practical and scalable strategy to foster psychological stability and cognitive resilience in expectant mothers.

Study Limitations

This study has several limitations that should be acknowledged. The cross-sectional design restricts causal inference, and recruitment from a single center in a socioeconomically low-to-middle region limits generalizability. Reliance on self-reported measures for cognitive flexibility, anxiety, and forgetfulness introduces potential response bias, while the inclusion of only one objective cognitive measure (verbal fluency) narrows the scope of cognitive evaluation. Unmeasured variables—such as sleep quality, nutritional status, hormonal fluctuations, and comorbid psychiatric conditions—may also have influenced outcomes. Furthermore, the modest sample size could limit statistical power, particularly in subgroup analyses.

Nevertheless, this study offers important strengths. It is among the few studies to integrate psychological and cognitive measures during pregnancy, using both subjective and objective assessments. The application of a validated pregnancy-specific anxiety scale ensured sensitivity to gestational concerns often overlooked by general anxiety tools. This multidimensional approach enhances both the methodological rigor and the clinical relevance of the findings. Future research should use longitudinal or interventional designs to clarify causality and test cognitive-flexibility–enhancing interventions. Wider executive and attentional assessments, more diverse samples, and stricter control of confounders will be critical for advancing this field.

Conclusion

Pregnant women in this study exhibited low-to-moderate levels of pregnancy-related anxiety and moderate-to-good cognitive flexibility. Greater cognitive flexibility was consistently associated with lower anxiety and better verbal fluency, underscoring its potential role as a protective cognitive resource during pregnancy. Although anxiety was not directly correlated with objective verbal fluency, self-reported attentional difficulties were associated with both higher anxiety and reduced cognitive flexibility, indicating a complex interplay among emotional state, subjective cognition, and performance-based outcomes. Cognitive flexibility, therefore, emerges as a promising target for interventions aimed at safeguarding maternal mental health and warrants further investigation in larger, more diverse, and methodologically rigorous studies.

Ethics

Ethics Committee Approval: The study protocol was approved by the Gaziosmanpaşa Training and Research Hospital Clinical Research Ethics Committee (approval number: 76, date: June 7, 2023). All procedures were carried out in accordance with the Declaration of Helsinki.

Informed Consent: Written informed consent was obtained from all participants.

Footnotes

Authorship Contributions: Surgical and Medical Practices - A.D., S.T.K., O.B.; Concept - A.D., S.T.K., O.B.; Design - A.D., S.T.K., O.B.; Data Collection or Processing - A.D., S.T.K.; Analysis or Interpretation - A.D., S.T.K., O.B.; Literature Search - A.D.; Writing - A.D., S.T.K., O.B.

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. Bjelica A, Cetkovic N, Trninic-Pjevic A, Mladenovic-Segedi L. The phenomenon of pregnancy - a psychological view. Ginekol Pol. 2018; 89: 102-6.
2. Eick SM, Goin DE, Izano MA, Cushing L, DeMicco E, Padula AM, et al. Relationships between psychosocial stressors among pregnant women in San Francisco: a path analysis. PLoS One. 2020; 15: e0234579.
3. Aziz HA, Yahya HDB, Ang WW, Lau Y. Global prevalence of depression, anxiety, and stress symptoms in different trimesters of pregnancy: a meta-analysis and meta-regression. J Psychiatr Res. 2025; 181: 528-46.
4. Grigoriadis S, Graves L, Peer M, Mamisashvili L, Tomlinson G, Vigod SN, et al. Maternal anxiety during pregnancy and the association with adverse perinatal outcomes: systematic review and meta-analysis. J Clin Psychiatry. 2018; 79: 17r12011.
5. Grattan DR, Ladyman SR. Neurophysiological and cognitive changes in pregnancy. Handb Clin Neurol. 2020; 171: 25-55.
6. Tong K, Chan YN, Cheng X, Cheon B, Ellefson M, Fauziana R, et al. Study protocol: how does cognitive flexibility relate to other executive functions and learning in healthy young adults? PLoS One. 2023; 18: e0286208.
7. Grant JE, Chamberlain SR. Impaired cognitive flexibility across psychiatric disorders. CNS Spectr. 2023; 28: 688-92.
8. Asali A, Miller N, Engel O, Haj-Yahia N, Cohen O, Ben-Kimhy R, et al. Does cognitive flexibility affect the course of labor? A prospective study. Int J Gynaecol Obstet. 2020; 148: 187-91.
9. McCormack C, Callaghan BL, Pawluski JL. It's time to rebrand "Mommy Brain". JAMA Neurol. 2023; 80: 335-6.
10. Younis J, Bleibel M, Masri JE, Ismail A, Abou-Abbas L. Exploring the influence of pregnancy on cognitive function in women: a systematic review. BMC Pregnancy Childbirth. 2025; 25: 88.
11. Albin-Brooks C, Nealer C, Sabihi S, Haim A, Leuner B. The influence of offspring, parity, and oxytocin on cognitive flexibility during the postpartum period. Horm Behav. 2017; 89: 130-6.
12. Brunton RJ, Dryer R, Saliba A, Kohlhoff J. Pregnancy anxiety: a systematic review of current scales. J Affect Disord. 2015; 176: 24-34.
13. Kurt G, Arslan H. Turkish version of the pregnancy-related anxiety scale: a psychometric study. Perspect Psychiatr Care. 2021; 57: 157-66.
14. Dennis JP, Vander Wal JS. The cognitive flexibility inventory: instrument development and estimates of reliability and validity. Cogn Ther Res. 2010; 34: 241-53.
15. Gulum IV, Dag I. Turkish adaptation, validity and reliability study of the Repetitive Thinking Questionnaire and the Cognitive Flexibility Inventory. Anadolu Psikiyatri Derg. 2012; 13: 216-23.
16. Lezak MD. Neuropsychological assessment. 2nd ed. New York (NY): Oxford University Press; 1983.
17. Vaughan RM, Coen RF, Kenny R, Lawlor BA. Semantic and phonemic verbal fluency discrepancy in mild cognitive impairment: potential predictor of progression to Alzheimer's disease. J Am Geriatr Soc. 2018; 66: 755-9.
18. Nyongesa P, Ekhaguere OA, Marete I, Tenge C, Kemoi M, Bann CM, et al. Maternal age extremes and adverse pregnancy outcomes in low-resourced settings. Front Glob Womens Health. 2023; 4: 1201037.
19. Silva MMJ, Nogueira DA, Clapis MJ, Leite EPRC. Anxiety in pregnancy: prevalence and associated factors. Rev Esc Enferm USP. 2017; 51: e03253.
20. Pinheiro RL, Areia AL, Mota Pinto A, Donato H. Advanced maternal age: adverse outcomes of pregnancy, a meta-analysis. Acta Med Port. 2019; 32: 219-26.
21. Demir A, Yazar M. Examination of stress, anxiety and depression during the prenatal period: a traditional review. KTO Karatay Univ Health Sci J. 2024; 4: 197-211.
22. Karabulutlu O, Yavuz C. Evaluation of the frequency of depression and anxiety levels in high-risk pregnancies. Caucasian J Sci. 2021; 8: 51-69.
23. Val A, Posse CM, Míguez MC. Risk factors for prenatal anxiety in European women: a review. J Clin Med. 2025; 14: 3248.
24. Soleimani R, Asgharnia M, Jalali MM, Eslamdoust-Siahestalkhi F, Kousha M, Aeinfar K. Comparison of cognitive function in different trimesters of pregnancy. Caspian J Neurol Sci. 2022; 8: 156-62.
25. Daalderop LA, Lagendijk J, Steegers EAP, El Marroun H, Posthumus AG. Psychological distress during pregnancy and adverse maternal and perinatal health outcomes: the role of socioeconomic status. Int J Gynaecol Obstet. 2023; 163: 920-30.
26. Bedaso A, Adams J, Peng W, Sibbritt D. The relationship between social support and mental health problems during pregnancy: a systematic review and meta-analysis. Reprod Health. 2021; 18: 162.
27. Maharramova A, Gur EY. The pregnant women's perception of risks and pregnancy stress levels: a cross-sectional study from Turkey. Rev Assoc Med Bras (1992). 2024; 70: e20231270.
28. Xu Y, Zheng P, Feng W, Chen L, Sun S, Liu J, et al. Patterns of attentional bias in antenatal depression: an eye-tracking study. Front Behav Neurosci. 2023; 17: 1288616.
29. Ali SA, Sualeh M, Raza G, Sabeeh Ul Haq M, Hissan L, Zafar D, et al. Exploring the impact of pregnancy on cognitive function: a comparative study in a low-income setting. BMC Pregnancy Childbirth. 2025; 25: 447.
30. Park J, Moghaddam B. Impact of anxiety on prefrontal cortex encoding of cognitive flexibility. Neuroscience. 2017; 345: 193-202.

İstanbul Med J 2026; 27(2): 136-42
DOI: 10.4274/imj.galenos.2026.59265

The Relationship Between Iron Stores, Type 2 Diabetes, and Diabetic Complications

Muhammet Ateş, Tayyibe Saler

University of Health Sciences Türkiye, Adana City Training and Research Hospital, Clinic of Internal Medicine, Adana, Türkiye

ABSTRACT

Introduction: The present investigation sought to examine the relationship of circulating ferritin concentrations with type 2 diabetes mellitus (T2DM) and with diabetes-related complications. Additionally, we aimed to determine whether serum ferritin could serve as a biomarker of metabolic and inflammatory status in individuals with diabetes.

Methods: This retrospective observational study included 422 subjects: 195 with T2DM, 82 with insulin resistance (IR), and 145 metabolically healthy controls. Demographic characteristics and biochemical variables were evaluated. Intergroup differences were analyzed using analysis of variance. Correlation analyses and receiver operating characteristic (ROC) curve assessments were performed to determine diagnostic accuracy. Furthermore, multivariable logistic regression was conducted to identify independent factors associated with diabetic complications.

Results: Participants with T2DM had significantly greater mean age and body mass index than control subjects ($p<0.001$). Serum ferritin concentrations were significantly increased in both the T2DM and IR groups relative to controls ($p<0.001$). Ferritin levels were positively associated with HbA1c (r: 0.179, $p=0.012$) and age (r: 0.133, $p=0.006$). In contrast, inverse relationships were observed between ferritin and both diabetes duration (r: −0.192, $p=0.007$) and initiation of insulin treatment (r: −0.243, $p=0.007$). The ROC curve analysis indicated that ferritin demonstrated fair discriminatory performance in identifying T2DM (area under the curve: 0.725, $p<0.001$). Elevated ferritin concentrations were observed in patients with diabetic neuropathy, retinopathy, and nephropathy compared with controls ($p<0.001$); however, no statistically significant differences were detected among the specific complication subtypes ($p=0.111$). In multivariate analysis, diabetes duration emerged as the sole independent determinant of diabetic complications ($p=0.046$).

Conclusion: Serum ferritin is significantly elevated in T2DM and correlates with poor glycemic control. Although ferritin reflects metabolic and inflammatory activity, diabetes duration remains the strongest predictor of diabetic complications.

Keywords: Type 2 diabetes mellitus, ferritin, insulin resistance, diabetic complications

Introduction

Type 2 diabetes mellitus (T2DM) represents a multifactorial metabolic disease defined by impaired insulin sensitivity, gradual deterioration of pancreatic ß-cell function, and persistent hyperglycemia (1). The global prevalence of T2DM has risen dramatically in recent decades, driven by changes in lifestyle, diet, and an aging population (2,3). Persistent hyperglycemia in diabetic individuals contributes to a wide range of complications, including cardiovascular disease, nephropathy, neuropathy, and retinopathy (4-6). Understanding the metabolic and biochemical pathways that contribute to the onset and progression of T2DM is therefore of major clinical and public health importance.

Iron metabolism has been increasingly recognized as a potential factor influencing glucose homeostasis and insulin sensitivity (7). Iron, while essential for oxygen transport and cellular energy metabolism, can exert toxic effects in excess through the generation of reactive oxygen species (8). Elevated body iron stores have been linked to oxidative stress, lipid peroxidation, and tissue damage—all of which may impair insulin action and pancreatic ß-cell function (9). Conversely, low iron levels may disrupt mitochondrial energy balance and metabolic regulation, suggesting that both iron overload and deficiency contribute to metabolic disorders.

Multiple investigations have reported a significant relationship between circulating ferritin levels—commonly regarded as an indirect indicator of total body iron stores—and the likelihood of developing T2DM (10-12). Elevated ferritin concentrations have consistently been documented in individuals with diabetes and are linked to insulin resistance (IR), systemic inflammatory activity, and features of the metabolic syndrome (MetS) (13). Moreover, abnormalities in iron

Address for Correspondence: Muhammet Ateş, MD, University of Health Sciences Türkiye, Adana City Training and Research Hospital, Clinic of Internal Medicine, Adana, Türkiye
E-mail: m_ates0@hotmail.com **ORCID ID:** orcid.org/0009-0009-8807-6010

Cite this article as: Ateş M, Saler T. The relationship between iron stores, type 2 diabetes, and diabetic complications. İstanbul Med J. 2026; 27(2): 136-42

Received: 19.02.2026
Accepted: 01.04.2026
Publication Date: 12.05.2026

homeostasis are thought to contribute to the development of diabetes-related complications, primarily via mechanisms involving endothelial dysfunction, increased oxidative stress, and microvascular impairment.

Beyond its function as an iron-storage protein, ferritin serves as an acute-phase reactant and may reflect the underlying low-grade inflammation frequently observed in metabolic disorders. Chronic subclinical inflammation is a recognized component of IR and type 2 diabetes, contributing to endothelial dysfunction and progressive microvascular injury. In this context, elevated ferritin concentrations may represent not only altered iron balance but also an integrated signal of inflammatory and metabolic stress. Distinguishing whether ferritin acts as a causal mediator or merely as a surrogate biomarker in the diabetic milieu remains an area of ongoing investigation, underscoring the need for further clinical and mechanistic studies.

The present study aims to investigate the relationship between iron stores and the presence of type 2 diabetes and its associated complications. By evaluating serum ferritin and related iron parameters in diabetic and non-diabetic individuals, this study seeks to clarify the potential role of iron metabolism in the pathophysiology of diabetes and its complications. A better understanding of these interactions may contribute to identifying novel biomarkers and therapeutic targets for preventing or mitigating diabetic complications.

Methods

Study Design and Population

This retrospective, observational study was conducted at the internal medicine outpatient clinics and inpatient wards of the Adana City Training And Research Hospital. Ethical approval was obtained from the Non-Interventional Clinical Research Ethics Committee of Çukurova University (meeting number: 66, date: July 7, 2017). Due to the retrospective nature of the study, the requirement for informed consent was waived by the ethics committee.

Medical records of adult patients (≥18 years) diagnosed with T2DM and/or IR between January 1, 2015 and July 1, 2017 were reviewed through Adana City Training and Research Hospital's electronic medical database. The study population included 422 participants: 195 patients with T2DM (123 receiving insulin and 72 receiving oral antidiabetic drugs), 82 individuals with IR, and 145 healthy controls who presented to the internal medicine clinics for routine check-ups.

Data Collection

Demographic characteristics and laboratory parameters were extracted from institutional medical records for all participants. The recorded variables included body mass index (BMI), fasting plasma glucose, glycated hemoglobin A1c (HbA1c), serum ferritin, low-density lipoprotein cholesterol, triglycerides, and serum creatinine levels. In individuals with diabetes, the presence of chronic microvascular complications—specifically retinopathy, nephropathy, and neuropathy—was documented, excluding diabetic foot involvement. BMI was computed as body weight in kilograms divided by the square of height measured in meters (kg/m^2). IR was quantified using the Homeostasis Model Assessment for IR (HOMA-IR), calculated using the following equation: HOMA-IR = (fasting insulin×fasting plasma glucose)/405. A threshold value of ≥2.5 was defined as indicative of IR. Serum ferritin concentrations were measured using an electrochemiluminescence immunoassay method with the Elecsys Ferritin assay kit (Roche Diagnostics, Mannheim, Germany) on the Elecsys 2010 analyzer platform. The established laboratory reference intervals were 30–400 ng/mL for males and 13–150 ng/mL for females.

Inclusion and Exclusion Criteria

This study included adult patients aged 18 years or older who had been diagnosed with T2DM, IR, or both and whose medical records were fully accessible in Adana City Training and Research Hospital's electronic database. Healthy individuals presenting to the internal medicine outpatient clinics for routine health examinations during the same period and with no known systemic or metabolic disorders were included as the control group. Patients were excluded from the study if they were younger than 18 years of age, were pregnant, or had incomplete or missing clinical data. Individuals with anemia (defined as hemoglobin <13 g/dL in men and <12 g/dL in women), with genetic disorders of iron metabolism (such as hemochromatosis or hemosiderosis), or with a history of iron supplementation or blood transfusion were also excluded. In addition, patients with chronic systemic diseases known to affect ferritin levels—such as chronic kidney disease, rheumatologic disorders, or malignancy—were not included. Those with diabetic foot or other active infectious diseases were likewise excluded from the analysis.

Statistical Analysis

Data were analyzed using SPSS software (version 27.0; IBM Corp., Chicago, IL, USA). Continuous variables are presented as mean ± standard deviation, whereas categorical data are summarized as counts and percentages. Normality of the distribution was assessed using the Kolmogorov–Smirnov test. For comparisons between two independent groups, the Student's t-test was applied when parametric assumptions were satisfied, and the Mann–Whitney U test was used for non-normally distributed variables. Differences among multiple groups were examined using one-way analysis of variance, followed by appropriate post hoc analyses when overall significance was observed. Equality of variances was evaluated with Levene's test, and corresponding F values were reported. Associations between continuous parameters, including serum ferritin, HbA1c, and diabetes duration, were examined using Pearson correlation analysis. Correlation strength was classified as weak (0.1–0.3), moderate (0.3–0.5), or strong (>0.5). All analyses were conducted using two-sided tests, and p values below 0.05 were considered statistically significant.

In addition, binary logistic regression analysis was performed to identify independent factors associated with diabetic complications and to estimate effect size using odds ratios (ORs) with 95% confidence intervals (CIs). Receiver operating characteristic (ROC) curve analysis was also conducted to evaluate the discriminatory performance of serum ferritin for predicting the presence of T2DM, and the area under the curve (AUC), sensitivity, specificity, and optimal cut-off value were calculated.

Results

The diabetes cohort was significantly older than the control group (55.8±11.6 vs. 39.2±15.5 years, p<0.001). BMI was likewise greater among patients with diabetes (30.5±5.3 kg/m^2) compared with controls (28.5±4.3 kg/m^2) (p<0.001). Mean serum ferritin concentrations were markedly higher in the diabetic group (96.7±82.5 ng/mL) than in the control group (45.5±33.7 ng/mL) (p<0.001). In contrast, the sex distribution did not differ significantly between groups (p=0.587) (Table 1).

Patients with diabetes were older than both the control and IR groups (55.8±11.6 vs. 39.2±15.5 and 39.9±12.3 years, respectively; p<0.001). The highest BMI values were observed in the IR group (33.4±6.2 kg/m^2), whereas the lowest values were recorded in controls (28.5±4.3 kg/m^2) (p<0.001). Ferritin concentrations were elevated in both the diabetes (96.7±82.5 ng/mL) and IR groups (87.6±66.4 ng/mL) compared with the control group (45.5±33.7 ng/mL), with statistical significance (p<0.001). The sex distribution did not differ among the three groups (p=0.622) (Table 2).

Serum ferritin demonstrated a weak but statistically significant positive association with age (r: 0.133, p=0.006). In contrast, no meaningful relationship was detected between ferritin and BMI (r: 0.010, p=0.842). An inverse correlation was identified between ferritin levels and diabetes duration (r: −0.192, p=0.007), and between ferritin levels and transition to insulin treatment (r: −0.243, p=0.007). Additionally, ferritin showed a modest positive correlation with HbA1c concentrations (r: 0.179, p=0.012) (Table 3).

Although ferritin levels were lower in patients with complications (88.1±66.8 ng/mL) compared to those without complications (100.5±88.5 ng/mL), the difference was not statistically significant (p=0.612). Similarly, the presence of nephropathy (p=0.309), retinopathy (p=0.937), or neuropathy (p=0.571) did not result in significant changes in ferritin levels. No significant differences in serum ferritin levels were observed either when all complications were evaluated together (p=0.583) or among patients with two or more complications (p=0.602) (Table 4).

The AUC for serum ferritin was 0.725, indicating that ferritin has moderate discriminative ability for predicting type 2 diabetes (p<0.001). Using a cut-off value of 48.5 ng/mL, the sensitivity and specificity were 72% and 70%, respectively. The 95% CI for the analysis was 0.672–0.778 (Table 5, Figure 1).

Among the variables included in the analysis, only the duration of type 2 diabetes showed a statistically significant association with the presence of complications (B: 0.121, p=0.046, OR: 1.129, 95% CI: 1.002–1.272). No significant associations were observed between complications and sex, age, BMI, transition to insulin therapy, serum ferritin levels or HbA1c levels (p>0.05) (Table 6).

Serum ferritin levels were significantly higher in patients with neuropathy (87.1±68.6 ng/mL), retinopathy (76.6±83.8 ng/mL), and nephropathy (110.2±75.1 ng/mL) compared to the control group (45.4±33.6 ng/mL) (p<0.001). However, when all complication types were evaluated together, no statistically significant differences were observed among them (p=0.111) (Table 7).

Discussion

In this study, serum ferritin levels were significantly higher in individuals with T2DM than in both the control and IR groups. The higher mean age and BMI observed in the diabetic group indicate a greater metabolic burden and a more pronounced systemic inflammatory state within this population. The positive association between ferritin and HbA1c, together with its negative relationships with diabetes duration and transition to insulin therapy, suggests that ferritin may act as a dynamic biomarker, reflecting acute metabolic activity and glycemic control, rather than as a static indicator of iron stores. Although ferritin levels tended to be higher in patients with diabetic complications, no

Table 1. Comparison of demographic and laboratory characteristics between control and diabetes groups

		Control (n=145)	Diabetes (n=195)	
Sex	Female	79 (41.4%)	112 (58.6%)	0.587
	Male	66 (44.3%)	83 (55.7%)	
Age (years)		39.2±15.5	55.8±11.6	**<0.001**
BMI (kg/m²)		28.5±4.3	30.5±5.3	**<0.001**
Serum ferritin (ng/mL)		45.5±33.7	96.7±82.5	**<0.001**

Values are expressed as mean ± standard deviation. BMI: Body mass index

Table 2. Comparison of demographic and laboratory characteristics among control, insulin resistance, and diabetes groups

		Control (n=145)	Insuline resistance (n=82)	Diabetes (n=195)	p value
Sex	Female	79 (33.9%)	42 (18.0%)	112 (48.1%)	0.622
	Male	66 (34.9%)	40 (21.2%)	83 (43.9%)	
Age (years)		39.2±15.5	39.9±12.3	55.8±11.6	**<0.001**
BMI (kg/m²)		28.5±4.3	33.4±6.2	30.5±5.3	**<0.001**
Serum Ferritin (ng/mL)		45.5±33.7	87.6±66.4	96.7±82.5	**<0.001**

Values are expressed as mean±standard deviation. Statistical significance post-hoc analysis. Age: control vs. diabetes: p<0.001, insulin resistance vs. diabetes: p<0.001. BMI: Control vs. insulin resistance: p<0.001, control vs. diabetes: p<0.01, insulin resistance vs. diabetes: p<0.001. Serum ferritin: Control vs. insulin resistance: p<0.001; control vs. diabetes: p<0.001. BMI: Body mass index

significant differences were observed across complication types. This pattern implies that ferritin may represent a marker of metabolic and inflammatory stress rather than directly mirroring the extent of diabetes-related organ damage. Collectively, these findings highlight the complex interplay between iron metabolism and glucose regulation and suggest that ferritin could serve as an indirect indicator of both metabolic control and systemic inflammation in patients with T2DM.

In the present analysis, serum ferritin concentrations were significantly higher in individuals with type 2 diabetes and IR compared with healthy controls. ROC curve analysis demonstrated that ferritin has moderate diagnostic performance in identifying T2DM (AUC: 0.725). At a threshold of 48.5 ng/mL, sensitivity was 72% and specificity was 70%.

Table 3. Correlations between serum ferritin levels and clinical parameters

		Serum ferritin (ng/mL)
Age (years)	r	0.133
	p	**0.006**
BMI (kg/m²)	r	0.010
	p	0.842
Duration of type 2 diabetes (years)	r	-0.192
	p	**0.007**
Transition to insulin therapy	r	-0.243
	p	**0.007**
Hemoglobin A1c (%)	Correlation coefficient (r)	0.179
	p value	**0.012**

r: Pearson correlation coefficient, BMI: Body mass index

Table 4. Comparison of serum ferritin levels according to the presence and type of complications in the diabetic population

		Mean ± SD	p value
Presence of complication	No	100.5±88.5	0.612
	Yes	88.1±66.8	
Diabetic nephropathy	No	95.5±83.4	0.309
	Yes	106.0±76.0	
Diabetic retinopathy	No	97.3±84.6	0.937
	Yes	93.9±72.5	
Diabetic neuropathy	No	99.1±85.6	0.571
	Yes	87.1±68.6	
All complication	No	97.7±83.5	0.583
	Yes	71.9±47.2	
2 or more complications	No	97.6±83.4	0.602
	Yes	85.1±71.2	

Values are expressed as mean ± standard deviation (ng/mL). SD: Standard deviation

These findings align with prior cross-sectional evidence indicating elevated ferritin levels in poorly controlled T2D and a parallel increase in ferritin with worsening glycemic status. Consistent with this, Bayih et al. (12) reported significantly greater ferritin levels in uncontrolled diabetic patients relative to both controlled diabetics and non-diabetic subjects, along with a positive association between ferritin and HbA1c. Genetic evidence from Mendelian randomization analyses further supports a causal contribution of iron status to T2DM risk. Wang et al. (14) demonstrated that genetically elevated systemic iron indices—including serum iron, ferritin, and transferrin saturation—were positively associated with T2DM risk. Similarly, a recent meta-analysis linking iron overload with MetS and metabolic dysfunction–associated steatotic liver disease established a dose–response relationship, highlighting ferritin as a validated biomarker of metabolic derangement, with a stronger effect observed among women (15). In line with these findings, observational and mechanistic studies have shown that ferritin rises in parallel with IR markers and inversely with HDL cholesterol, supporting the directionality of our results in both T2DM and IR populations (13).

In our dataset, ferritin showed a significant positive correlation with HbA1c and negative associations with diabetes duration and transition to insulin therapy. The positive relationship with HbA1c reinforces the concept that ferritin may reflect long-term glycemic burden, a finding consistent with the study by Bayih et al. (12), who reported a moderate-to-strong ferritin–HbA1c correlation among individuals with T2DM. Moreover, the link between ferritin and the IR phenotype has been confirmed on a population scale using the MetS-IR index; a large NHANES-based analysis identified an independent and linear association between ferritin and MetS-IR, particularly among women (16). The absence of a significant relationship between ferritin

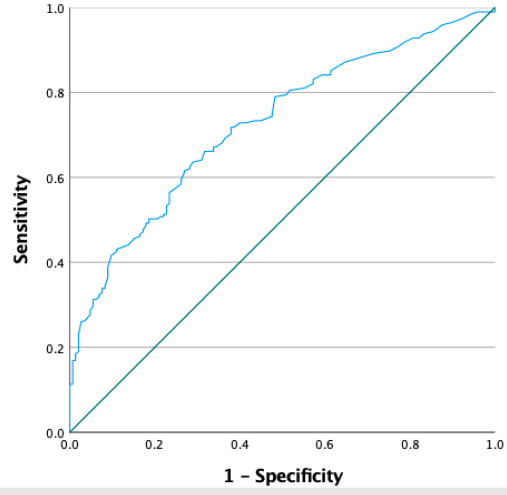

Figure 1. Receiver operating characteristic curve of serum ferritin levels for predicting type 2 diabetes

Table 5. ROC analysis for serum ferritin in predicting the presence of type 2 diabetes

	Area	Cut-off	Sensitivity	Specificity	p value	Asymptotic 95% CI	
						Lower bound	Upper bound
Ferritin	0.725	48.5	72	70	**<0.001**	0.672	0.778

ROC: Receiver operating characteristic, CI: Confidence interval

Table 6. Binary logistic regression analysis for predictors of diabetic complications

	B	S.E.	Sig.	Exp (B)	95% CI for exp. (B)
Sex (female)	0.169	0.457	0.712	1.184	0.483–2.899
Age (years)	0.013	0.022	0.553	1.013	0.971–1.056
BMI (kg/m²)	0.017	0.043	0.694	1.017	0.935–1.105
Duration of type 2 DM (years)	**0.121**	**0.061**	**0.046**	**1.129**	**1.002–1.272**
Transition to insulin therapy	0.042	0.070	0.545	1.043	0.910–1.197
Serum ferritin (ng/mL)	–0.002	0.003	0.562	0.998	0.993–1.004
Hemoglobin A1c (%)	0.119	0.116	0.307	1.126	0.896–1.415
Constant	–4.226	2.507	0.092	0.015	–

B: Regression coefficient, S.E.: Standard error, CI: Confidence interval, DM: Diabetes mellitus, BMI: Body mass index, exp.: Exponential, Sig: Significance

Table 7. Comparison of serum ferritin levels among diabetic complication types

Complication type	Group	Ferritin (ng/mL)	p value
Neuropathy	Neuropathy	87.1±68.6	**<0.001**
	Uncomplicated DM	96.9±85.8	
	Control	45.4±33.6	
Retinopathy	Retinopathy	76.6±83.8	**<0.001**
	Uncomplicated DM	96.9±85.8	
	Control	45.4±33.6	
Nephropathy	Nephropathy	110.2±75.1	**<0.001**
	Uncomplicated DM	96.9±85.8	
	Control	45.4±33.6	
All complications	Neuropathy	89.7+70.5	**0.111**
	Nephropathy	110.2+75.1	
	Retinopathy	60.2+56.9	

Values are expressed as mean ± standard deviation (ng/mL). DM: Diabetes mellitus

and BMI in our study suggests that confounding factors such as visceral adiposity and low-grade inflammation may influence this association. This interpretation aligns with a recent meta-analysis indicating that the ferritin– MetS relationship may be confounded by both BMI and C-reactive protein (15). In addition, contemporary reviews summarizing the iron–insulin axis and the bidirectional interaction between hepcidin-mediated inflammation and IR further clarify the biological basis of the observed ferritin–glycemic linkage (17).

In our study, serum ferritin levels were significantly higher in diabetic subgroups with neuropathy, retinopathy, and nephropathy than in the control group, suggesting a potential association between iron overload and diabetes-related organ damage. However, no significant difference in ferritin concentrations was observed among the different complication types, indicating that ferritin may reflect systemic metabolic stress rather than the specific pattern of end-organ involvement. Experimental data have shown that retinal ferritin expression increases in response to systemic iron accumulation, yet the penetration of circulating iron into retinal tissues appears limited, implying that serum ferritin may serve only as an indirect or confounded marker for diabetic retinopathy (18). Clinically, studies have also reported that **serum iron**, rather than ferritin, may be inversely related to the presence of retinopathy, emphasizing the heterogeneity of the iron–retinopathy relationship (19). Regarding renal involvement, recent reviews have highlighted the critical role of ferroptosis and oxidative stress pathways in the pathogenesis of diabetic kidney disease (DKD), suggesting that excess iron may aggravate tubulointerstitial injury, thereby explaining the relatively higher ferritin levels observed in our nephropathy subgroup (20). Emerging evidence on urinary ferritin as a potential early biomarker for DKD further supports the notion that circulating ferritin alone may not adequately capture the severity or progression of diabetic complications (21).

In multivariate logistic regression analysis, only the duration of type 2 diabetes was independently associated with diabetic complications, whereas serum ferritin did not emerge as a significant predictor. The predominance of disease duration and chronic glycemic exposure as determinants of microvascular damage has been consistently demonstrated in large-scale population studies; for instance, studies conducted in Chinese adults identified diabetes duration and glycemic control as the principal risk factors for diabetic retinopathy (22). Ferritin's limited predictive performance may be partly attributed to its nature as an acute-phase reactant and to its modulation by hepcidin-dependent pathways, which can attenuate its independent association in adjusted models. Indeed, in patients with T2DM, hepcidin levels, rather than ferritin, have been shown to correlate with long-term fatal and non-fatal cardiovascular outcomes, highlighting the stronger prognostic role of regulatory hormones within iron metabolism (23). Collectively, these findings suggest that while ferritin reflects aspects of metabolic and inflammatory burden, disease duration and glycemic indicators remain the primary determinants in predicting diabetic complications and should be prioritized in clinical risk assessment.

The effects of iron on ß-cell function, hepatic gluconeogenesis, adipocyte signaling, and oxidative stress underscore its central role in the pathophysiology of T2DM. A recent review highlighted that iron overload accelerates ß-cell dysfunction through ferroptosis, enhances lipid peroxidation, and contributes to worsening glycemic control (24). On the other hand, hepcidin, the key hormonal regulator of iron homeostasis, is also upregulated by inflammation and obesity via the JAK–STAT3 and BMP–SMAD signaling pathways. Therefore, failure to interpret hepcidin levels in conjunction with ferritin may lead to misclassification of metabolic iron status. In this context, comprehensive analyses of systemic iron signaling and insulin sensitivity have emphasized the need for caution when interpreting ferritin-based risk assessments (17). Finally, genetic evidence from

Mendelian randomization studies supports a causal contribution of systemic iron status to T2DM risk, reinforcing the potential benefit of therapeutic strategies that reduce iron load, such as targeted phlebotomy, iron-restricted diets, or modulation of the hepcidin–ferroportin axis (14). This causal inference aligns with our finding of elevated ferritin levels in both the T2DM and IR groups, further supporting the biological plausibility of our results.

Study Limitations

This study has several limitations. First, it was designed as a retrospective, single-center study, which may limit the generalizability of the findings due to potential selection bias and the characteristics of the local patient population. Factors that could influence serum ferritin levels—such as acute-phase reactions, subclinical inflammation, or mild hepatic dysfunction—were not assessed; therefore, ferritin may reflect not only iron storage status but also an underlying inflammatory response. Additionally, other markers of iron metabolism (e.g., serum iron, transferrin saturation, hepcidin, or soluble transferrin receptor) were not evaluated, limiting the ability to comprehensively define ferritin's pathophysiological role within iron homeostasis. The presence of diabetic complications was ascertained from medical records, which may introduce diagnostic misclassification bias. Furthermore, the relatively modest sample size and the cross-sectional nature of the data prevented the evaluation of temporal changes in ferritin levels and their potential causal relationship with the development of diabetic complications. Despite these limitations, this study provides valuable insights by simultaneously assessing diabetes, IR, and ferritin levels within the same population and exploring the relationship between ferritin and different diabetic complications, thereby contributing meaningful data to the existing literature.

Conclusion

Serum ferritin concentrations were significantly higher in individuals with T2DM and IR than in non-diabetic controls. Ferritin showed moderate accuracy in distinguishing T2DM and demonstrated a positive relationship with HbA1c, indicating its association with metabolic and inflammatory load. Although ferritin levels were elevated in patients with microvascular complications compared with controls, they did not independently predict microvascular complications after multivariable adjustment. Disease duration was the only independent factor linked to diabetic complications. Overall, ferritin appears to reflect metabolic dysregulation rather than serving as a reliable standalone predictor of diabetes-related organ damage. Large-scale prospective studies incorporating broader iron metabolism parameters are needed to better define its clinical relevance in risk assessment and management of type 2 diabetes.

Ethics

Ethics Committee Approval: Ethical approval was obtained from the Non-Interventional Clinical Research Ethics Committee of Çukurova University (meeting number: 66, date: July 7, 2017).

Informed Consent: Due to the retrospective nature of the study, the requirement for informed consent was waived by the ethics committee.

Footnotes

Authorship Contributions: Surgical and Medical Practices - M.A.; Concept - T.S.; Design - T.S.; Data Collection or Processing - M.A.; Analysis or Interpretation - T.S.; Literature Search - M.A.; Writing - M.A.

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. Singh A, Shadangi S, Gupta PK, Rana S. Type 2 diabetes mellitus: a comprehensive review of pathophysiology, comorbidities, and emerging therapies. Compr Physiol. 2025; 15: e70003.
2. Yan Y, Wu T, Zhang M, Li C, Liu Q, Li F. Prevalence, awareness and control of type 2 diabetes mellitus and risk factors in Chinese elderly population. BMC Public Health. 2022; 22: 1382.
3. Kalan Farmanfarma KH, Ansari-Moghaddam A, Zareban I, Adineh HA. Prevalence of type 2 diabetes in Middle-East: systematic review & meta-analysis. Prim Care Diabetes. 2020; 14: 297-304.
4. van Sloten TT, Sedaghat S, Carnethon MR, Launer LJ, Stehouwer CDA. Cerebral microvascular complications of type 2 diabetes: stroke, cognitive dysfunction, and depression. Lancet Diabetes Endocrinol. 2020; 8: 325-36.
5. Viigimaa M, Sachinidis A, Toumpourleka M, Koutsampasopoulos K, Alliksoo S, Titma T. Macrovascular complications of type 2 diabetes mellitus. Curr Vasc Pharmacol. 2020; 18: 110-6.
6. Lyssenko V, Vaag A. Genetics of diabetes-associated microvascular complications. Diabetologia. 2023; 66: 1601-13.
7. Kirkutyte I, Latunde-Dada GO. Relationship between hepatic iron concentration and glycemic metabolism, prediabetes, and type 2 diabetes: a systematic review. Nutr Rev. 2025; 83: e1828-42.
8. Kawabata T. Iron-induced oxidative stress in human diseases. Cells. 2022; 11: 2152.
9. Marku A, Galli A, Marciani P, Dule N, Perego C, Castagna M. Iron metabolism in pancreatic beta-cell function and dysfunction. Cells. 2021; 10: 2841.
10. Conway RB, Pratte KA, Bowler RP, Young KA, Kinney GL, Austin E, et al. Plasma proteomic markers of iron and risk of diabetes in a Cohort of African American and White American current and former smokers. Diabetes Metab Syndr Obes. 2024; 17: 4767-76.
11. Gallardo-Nuell L, Blanch J, Leal Y, Coral DE, Duarte-Salles T, Giordano GN, et al. BMI-residualized data uncovers a cluster of people with type 2 diabetes and increased serum ferritin protected from cardiovascular disease. Cardiovasc Diabetol. 2025; 24: 139.
12. Bayih A, Dedefo G, Kinde S, Alem M, Negesso AE, Baye A, et al. Serum ferritin level and associated factors among uncontrolled adult type II diabetic follow-up patients: comparative based cross-sectional study. BMC Endocr Disord. 2024; 24: 144.
13. Srivastav SK, Mir IA, Bansal N, Singh PK, Kumari R, Deshmukh A. Serum ferritin in metabolic syndrome-mechanisms and clinical applications. pathophysiology. 2022; 29: 319-25.
14. Wang X, Fang X, Zheng W, Zhou J, Song Z, Xu M, et al. Genetic support of a causal relationship between iron status and type 2 diabetes: a Mendelian randomization study. J Clin Endocrinol Metab. 2021; 106: e4641-51.

15. Yu L, Que T, Zhou Y, Liu Z. Dose-response relationship of serum ferritin and dietary iron intake with metabolic syndrome and non-alcoholic fatty liver disease incidence: a systematic review and meta-analysis. Front Nutr. 2024; 11: 1437681.

16. Hao H, Chen Y, Xiaojuan J, Siqi Z, Hailiang C, Xiaoxing S, et al. The association between METS-IR and serum ferritin level in United States female: a cross-sectional study based on NHANES. Front Med (Lausanne). 2022; 9: 925344.

17. Sobieska K, Buczyńska A, Krętowski AJ, Popławska-Kita A. Iron homeostasis and insulin sensitivity: unraveling the complex interactions. Rev Endocr Metab Disord. 2024; 25: 925-39.

18. Bonet A, Pampalona J, Jose-Cunilleras E, Nacher V, Ruberte J. Ferritin but not iron increases in retina upon systemic iron overload in diabetic and iron-dextran injected mice. Invest Ophthalmol Vis Sci. 2023; 64: 22.

19. Chen YJ, Chen JT, Tai MC, Liang CM, Chen YY, Chen WL. Serum iron and risk of diabetic retinopathy. Nutrients. 2020; 12: 2297.

20. Chu J, Wang K, Lu L, Zhao H, Hu J, Xiao W, et al. Advances of iron and ferroptosis in diabetic kidney disease. Kidney Int Rep. 2024; 9: 1972-85.

21. Huang B, Ding X, Wen W, Ye S. Urinary ferritin as an early indicator of tubular injury in diabetic kidney disease: insights from NHANES and clinical cohorts. Diabetes Metab Syndr Obes. 2025; 18: 2429-38.

22. Ma Y, Wang H, Jiang J, Han C, Lu C, Zeng S, et al. Prevalence of and risk factors for diabetic retinopathy in residents with different types of abnormal glucose metabolism with or without hypertension: a suburban community-based cross-sectional study. Front Endocrinol (Lausanne). 2022; 13:966619.

23. Mantovani A, Busti F, Borella N, Scoccia E, Pecoraro B, Sani E, et al. Elevated plasma hepcidin concentrations are associated with an increased risk of mortality and nonfatal cardiovascular events in patients with type 2 diabetes: a prospective study. Cardiovasc Diabetol. 2024; 23: 305.

24. Miao R, Fang X, Zhang Y, Wei J, Zhang Y, Tian J. Iron metabolism and ferroptosis in type 2 diabetes mellitus and complications: mechanisms and therapeutic opportunities. Cell Death Dis. 2023; 14: 186.

İstanbul Med J 2026; 27(2): 143-8
DOI: 10.4274/imj.galenos.2026.95826

Evaluation of Artificial Intelligence-Generated Information About Abdominal Ultrasonography

Ali Salbas[1], Gözde Merve Tekel[2], Aslı Dilara Büyüktoka[1], Raşit Eren Büyüktoka[3], Ali Murat Koc[1], Atilla Hikmet Çilengir[2]

[1]İzmir Katip Çelebi University, Atatürk Training and Research Hospital, Department of Radiology, İzmir, Türkiye
[2]İzmir Democracy University, Buca Seyfi Demirsoy Training and Research Hospital, Department of Radiology, İzmir, Türkiye
[3]University of Health Sciences Türkiye, İzmir City Hospital, Department of Radiology, İzmir, Türkiye

ABSTRACT

Introduction: This study examined the relevance, accuracy, clarity, and completeness of ChatGPT-5 responses to frequently asked patient questions about abdominal ultrasonography and considered the potential role of large language models (LLMs) as supportive tools in patient education.

Methods: This cross-sectional study analyzed ChatGPT-5 responses to 15 frequently asked questions from patients about abdominal ultrasonography. The questions were collected from Google's "other questions" section. Each question was entered into ChatGPT-5 in a separate session, and the model's answers were recorded. Ten radiologists independently evaluated the responses using four criteria: relevance, accuracy, clarity, and completeness, with each criterion scored on a 1-to-5 scale. Interrater reliability was assessed using the intraclass correlation coefficient (ICC).

Results: ChatGPT-5 demonstrated high performance across all evaluated criteria. Mean scores were 4.97±0.18 for relevance, 4.78±0.49 for accuracy, 4.85±0.40 for clarity, and 4.68±0.53 for completeness, with an overall mean of 4.82±0.26. The minimum score assigned by the evaluators was 3. ICC values were 0.266 for relevance, 0.236 for accuracy, 0.230 for clarity, 0.582 for completeness, and 0.555 for the total score.

Conclusion: ChatGPT-5 provided generally well-rated responses to common patient questions about abdominal ultrasonography. Although interrater reliability showed variable levels of agreement, moderate agreement was observed for completeness and total scores. The model's overall performance was favorable, suggesting that LLMs may function as supportive resources for patient education. Their use should remain complementary to professional medical guidance. Further studies with broader question sets, diverse patient populations, and multiple language models are warranted.

Keywords: Artificial intelligence, health information technology, natural language processing, patient communication, ultrasonography, abdominal

Introduction

Large language models (LLMs) are artificial intelligence systems trained on extensive text datasets that can generate human-like responses and interact with users in natural language (1). Recent studies have examined LLMs in many medical specialties, exploring their use in clinical and educational settings (2). Within radiology, they have been studied for applications such as diagnostic support, report generation, and educational assessment in radiology (3,4). Additionally, LLMs are being studied as tools to help improve communication between clinicians and patients (5,6).

Ultrasonography is one of the most widely used imaging modalities in daily clinical practice because it is inexpensive, easily accessible, and free of ionizing radiation. In Türkiye, ultrasonography represents the highest imaging volume nationwide, exceeding 30 million examinations in 2020 and 35 million in 2021 according to national health statistics (7). Before undergoing imaging, many patients seek information about the procedure and often express concerns related to preparation, comfort, or diagnostic value (8).

In recent years, patients have increasingly relied on internet-based resources to obtain health-related information (9). This trend has contributed to the rising use of artificial intelligence chatbots. Although these models can provide rapid and structured answers, the accuracy and reliability of the information they offer remain important questions (10).

Address for Correspondence: Ali Salbas, MD, İzmir Katip Çelebi University, Atatürk Training and Research Hospital, Department of Radiology, İzmir, Türkiye
E-mail: dralisalbas@gmail.com **ORCID ID:** orcid.org/0000-0002-6157-6367

Cite this article as: Salbas A, Tekel GM, Büyüktoka AD, Büyüktoka RE, Koc AM, Çilengir AH. Evaluation of artificial intelligence-generated information about abdominal ultrasonography. İstanbul Med J. 2026; 27(2): 143-8

Received: 18.02.2026
Accepted: 07.04.2026
Publication Date: 12.05.2026

The potential of LLMs to support patient education has been investigated in radiology and in other medical fields (11,12). However, the quality of LLM responses to frequently asked patient questions about ultrasonography in general and abdominal ultrasonography in particular has not been systematically evaluated to date. To our knowledge, no study has evaluated how ChatGPT-5 performs in this context. The purpose of this study is to assess the relevance, accuracy, clarity, and completeness of ChatGPT-5's answers to frequently asked patient questions regarding abdominal ultrasonography, and to discuss its potential role as a supportive tool in patient education.

Methods

Ethical approval for this study was obtained from the İzmir Katip Çelebi University Health Research Ethics Committee (decision number: 0670, date: 06.11.2025). Due to the use of publicly available data and the absence of direct human participant involvement or identifiable patient information, the requirement for informed consent was waived by the ethics committee. This cross-sectional study was designed to assess the quality of responses generated by ChatGPT-5 to patients' frequently asked questions about abdominal ultrasonography. To obtain the patient questions, a new Google (Alphabet Inc., Mountain View, CA, USA) account without any prior search history was used (13). The phrase "frequently asked questions about abdominal ultrasonography" was entered into Google using its Turkish equivalent. This search yielded 150 questions listed in the "other questions" section of the results page. Two board-certified radiologists with 8 and 10 years of experience in ultrasonography independently reviewed all items. After eliminating duplicates and questions with overlapping meanings, they reached a consensus on a final set of 15 unique, patient-oriented questions through joint discussion (Table 1, Figure 1). Each question was preserved in its original form as displayed in the Google results to maintain authentic patient language.

The selected questions were entered into ChatGPT-5 (OpenAI Inc., San Francisco, CA, USA) using an account with no prior conversations. The model was accessed through its official web interface, which offers three modes of operation: Auto, Instant, and Thinking (14). In this study, the Auto mode was used because it reflects the default setting and typical real-world user interaction with the model. No model parameters, including the temperature (which is fixed in the web interface), were adjusted. No system-level prompts or hidden instructions were modified, and all questions were submitted in their original form without any additional prompting or formatting. No additional prompts or contextual information was provided. Each question was submitted separately and asked only once. To prevent any influence from previous interactions, a new session was opened for every question by clearing the chat history (15,16). All queries were entered on November 18, 2025.

The responses produced by ChatGPT-5 were evaluated by 10 radiologists from six different institutions, each with 4 to 10 years of experience in abdominal ultrasonography. Each response was assessed using four criteria: relevance, accuracy, clarity, and completeness. Scores ranged from 1 to 5 for each criterion, where 1 indicated the lowest score and 5 indicated the highest score. The scoring framework was adapted from previously published studies (5,6). Relevance referred to how directly the answer addressed the question. Accuracy reflected whether the information was medically accurate and clinically appropriate. Clarity indicated how understandable and well-organized the response was. Completeness assessed whether the answer provided sufficient information to fully address the question. The evaluators were blinded to the source of the responses and informed only that they were reviewing answers to patients' questions about abdominal ultrasonography. They were not informed that the responses were generated by a LLM. All questions were submitted in Turkish, and the radiologists evaluated the responses in Turkish.

Statistical Analysis

All data were analyzed using IBM SPSS Statistics version 26.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics were reported as means and standard deviations for each evaluation criterion. Interrater reliability

Table 1. The 15 most frequently asked patient questions about abdominal ultrasonography that were submitted to ChatGPT-5

No	Question
1	What diseases can be detected on an abdominal ultrasound?
2	What should be done before an abdominal ultrasound?
3	What is examined with an abdominal ultrasound?
4	Can intestinal problems be seen on ultrasound?
5	How long does an abdominal ultrasound take?
6	Can you drink coffee before an abdominal ultrasound?
7	How much water should be drunk before an ultrasound?
8	Do you need to remove clothing during an ultrasound?
9	What happens if I do not drink water before the ultrasound?
10	Is the uterus visible on an abdominal ultrasound?
11	Does it matter if you are fasting or not for an ultrasound?
12	Can cancer be detected on an abdominal ultrasound?
13	Can you smoke before having an ultrasound?
14	Is an abdominal ultrasound harmful?
15	Why is an abdominal ultrasound ordered?

Note: Although shown in English for readability, all questions were originally submitted to ChatGPT-5 in Turkish

was assessed using the intraclass correlation coefficient (ICC), based on a two-way random-effects model with absolute agreement [ICC (2,k)], where k represents the number of raters. Ninety-five percent confidence intervals and p values were calculated for all ICC estimates. A p value <0.05 was considered statistically significant.

Results

Fifteen questions were assessed by 10 radiologists using four predefined criteria: relevance, accuracy, clarity, and completeness. All ratings were based on a 5-point scale. The lowest score assigned by the evaluators to any response was 3 out of 5 in each subcategory of the rating scale. Relevance had the highest overall mean score with a value of 4.97±0.18. The mean accuracy score was 4.78±0.49, while clarity had a mean score of 4.85±0.40. Completeness had the lowest mean among the four criteria with a value of 4.68±0.53. The total score calculated by averaging the four criteria for each evaluation was 4.82±0.26. Question-based descriptive values for all criteria are presented in Table 2. The distributions of mean scores and standard deviations for all four criteria are shown in Figure 2.

The interrater reliability analysis demonstrated varying levels of agreement among evaluators. ICC values were 0.266 for relevance, 0.236 for accuracy, 0.230 for clarity, 0.582 for completeness, and 0.555 for the total scores (Table 3). Completeness demonstrated the highest ICC value, followed by the total score; both values indicate moderate agreement, whereas relevance, accuracy, and clarity showed lower levels of agreement. Statistical significance was observed for completeness (p=0.001) and the total score (p=0.004). Relevance (p=0.144), accuracy (p=0.181), and clarity (p=0.194) did not reach statistical significance.

Discussion

This study evaluated the quality of ChatGPT-5 responses to frequently asked patient questions about abdominal ultrasonography. The analysis showed that ChatGPT-5 provided responses with favorable scores across all four evaluation criteria. Mean scores ranged from 4.68 to 4.97 on a 5-point scale, with an overall average of 4.82. Relevance received the highest scores, while completeness received the lowest scores. Interrater reliability showed variable levels of agreement, with ICC values ranging from 0.230 to 0.582. Agreement was higher for completeness and the total score, and lower for relevance, accuracy, and clarity. To our knowledge, this is the first study to evaluate LLM responses to patient questions specifically about abdominal ultrasonography. These findings suggest that while ChatGPT-5 can generate well-rated responses, variability remains in how medical experts evaluate these answers.

LLMs have increasingly been examined as tools for patient communication across different medical fields (17). A recent study comparing multiple LLMs for computed tomography (CT) and magnetic resonance imaging (MRI)-related patient questions found

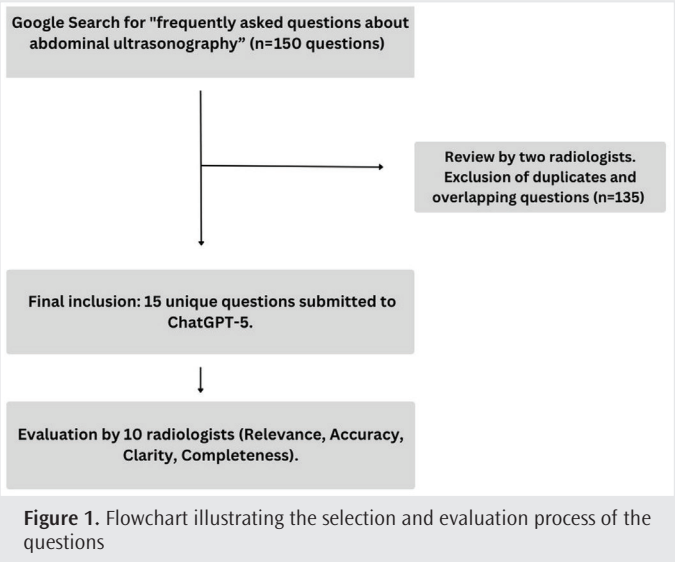

Figure 1. Flowchart illustrating the selection and evaluation process of the questions

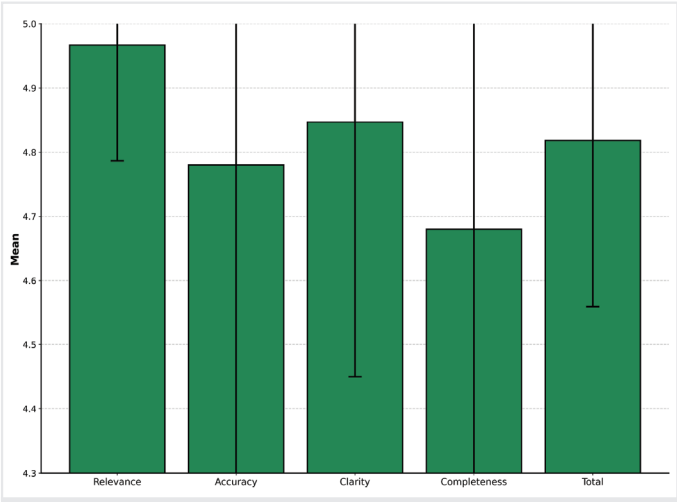

Figure 2. Mean scores and standard deviations for relevance, accuracy, clarity, completeness, and total scores

Table 2. Descriptive evaluation scores for each question across four assessment criteria

No	Relevance	Accuracy	Clarity	Completeness
1	4.9±0.32	4.7±0.48	4.8±0.42	4.3±0.67
2	5.0±0.00	4.8±0.42	4.9±0.32	4.9±0.32
3	5.0±0.00	5.0±0.00	4.9±0.32	4.3±0.67
4	4.8±0.42	4.4±0.70	4.5±0.85	4.3±0.82
5	5.0±0.00	4.6±0.84	4.9±0.32	4.7±0.48
6	5.0±0.00	4.6±0.70	5.0±0.00	4.7±0.48
7	4.9±0.32	4.8±0.42	4.9±0.32	4.8±0.42
8	5.0±0.00	4.8±0.42	4.9±0.32	4.5±0.53
9	5.0±0.00	4.9±0.32	5.0±0.00	4.9±0.32
10	4.9±0.32	4.9±0.32	4.6±0.52	4.9±0.32
11	5.0±0.00	5.0±0.00	4.9±0.32	4.8±0.42
12	5.0±0.00	4.7±0.67	4.9±0.32	4.5±0.71
13	5.0±0.00	4.9±0.32	4.9±0.32	4.9±0.32
14	5.0±0.00	4.9±0.32	4.9±0.32	4.9±0.32
15	4.9±0.32	4.9±0.32	4.9±0.32	4.8±0.42
Total	4.97±0.18	4.78±0.49	4.85±0.40	4.68±0.53

Values are presented as mean ± standard deviation. Each value represents the average score of 10 radiologists who evaluated the answers to 15 patient questions across four criteria: relevance, accuracy, clarity, and completeness

Table 3. ICCs, confidence intervals, and p values for all evaluation criteria

	ICC	Lower bound (95% CI)	Upper bound (95% CI)	p
Relevance	0.266	-0.220	0.730	0.144
Accuracy	0.236	-0.350	0.680	0.181
Clarity	0.230	-0.390	0.680	0.194
Completeness	0.582	0.240	0.830	0.001
Total	0.555	0.220	0.822	0.004

CI: Confidence interval, ICC: Intraclass correlation coefficient. ICC (2,k) model with absolute agreement was used, where k=10 raters.

that ChatGPT-4o achieved the highest scores for CT questions (mean 4.52) and shared the top position with Claude 3.5 Sonnet for MRI questions (both 4.79) (18). In cardiac imaging questions, ChatGPT-4 demonstrated 78.3% accuracy, with 86.7% of responses rated as clear and 81.7% as comprehensive (19). Similarly, Gordon et al. (20) found that 83% of ChatGPT responses to patient questions about radiological imaging were accurate, with 99% being at least partially relevant. However, performance varies with procedural complexity. While ChatGPT-3.5 adequately conveyed basic information for interventional radiology patient brochures, errors were infrequent for simpler procedures such as breast biopsy, whereas significant issues were noted for more complex topics including lung ablation and transarterial radioembolization (21). On questions about obstetric ultrasonography, both ChatGPT-3.5 and ChatGPT-4.0 answered 95% of questions correctly, significantly outperforming Microsoft Copilot (22). When evaluating responses to thyroid nodule questions, Campbell et al. (23) found 69.2% accuracy, though hallucination issues emerged when the model generated references.

Beyond radiology, studies in other specialties have also evaluated LLM performance. Physicians from 17 specialties rated ChatGPT responses with a median accuracy of 5.5 on a 6-point scale (24). In inflammatory bowel disease, ChatGPT responses showed no significant difference from physician responses in overall quality and accuracy, with superior completeness (25). For questions about anterior cruciate ligament surgery, both ChatGPT-4o and DeepSeek R1 demonstrated high accuracy with mean scores of 3.9 out of 4 (26). In contrast, breast reconstruction materials generated by ChatGPT-3.5 showed only 50% accuracy compared to expert-written content (27). In our study, ChatGPT-5 achieved an overall mean score of 4.82 across all evaluation criteria for abdominal ultrasonography questions, and scores for individual criteria ranged from 4.68 to 4.97. These findings are broadly consistent with previous research showing that LLMs generally perform well in answering patient-focused imaging questions. However, their reliability may vary depending on the clinical domain and the complexity of the information being evaluated.

The responses generated by ChatGPT-5 for patient questions about abdominal ultrasonography received generally high scores in this study. However, interrater agreement varied across the evaluation criteria and was higher for completeness and the total score but lower for relevance, accuracy, and clarity. This pattern has also been observed in studies evaluating LLMs across different medical fields. Hofmann et al. (28), who examined informed consent materials in interventional radiology, reported weak interrater agreement despite the high quality scores achieved by ChatGPT-4. Similarly, a study evaluating ChatGPT responses about celiac disease reported low reliability between two specialists (29). Studies in the field of orthopedics have also documented low agreement among evaluators, further supporting this observation (5,6). These findings suggest that variability in experts' evaluations may still occur when assessing text generated by LLMs, even when overall response quality is high.

Although the evaluators used the same scoring scales, aspects such as the perceived level of detail, tone of the responses, or subtle nuances in the content may have been interpreted differently. This should not be viewed as an error on the part of the evaluators. The ICC quantifies the proportion of total variance attributable to shared variance. For the ICC to yield a meaningful result, there must be sufficient variability in the scores (30). In this study, variability was limited for some of the evaluation criteria. For example, the mean relevance score of 4.97 indicates that the radiologists assigned a score of 5 to most of the questions. This limited variability may have influenced the ICC estimates, particularly for relevance, accuracy, and clarity, for which agreement remained lower despite generally high scores.

The high scores achieved by ChatGPT-5 in this study suggest that LLMs may serve as supportive tools for developing patient-education materials. Even so, the general limitations of these systems should be taken into account. Their training data may not fully reflect the most recent developments in medical practice, and they may occasionally generate information that is plausible yet incorrect (31). For this reason, model-generated responses cannot substitute for professional medical advice or individualized clinical judgment. Such tools are best used under expert oversight, particularly in settings where patients may rely heavily on the accuracy and clarity of the information provided. Given the widespread use of abdominal ultrasonography in routine clinical practice, a reliable and understandable supplementary resource may support patients' health literacy.

Study Limitations

This study has several limitations. The most important limitation is the lack of a patient perspective, as all evaluations were conducted solely by radiologists. In addition, no standardized readability or patient-oriented assessment tools were used, which may limit assessment of clarity from a patient perspective. A second limitation is that only a single LLM was examined, and the findings may differ when other models are included. Third, all patient questions were gathered from a single source, and broader sampling from platforms such as social media or patient forums might have provided greater diversity. In addition, the study included only Turkish-language questions, relied on responses collected at a single time point, and used a limited set of fifteen questions. Furthermore, each question was submitted only once, and potential variability in responses across repeated queries was not assessed. Given the stochastic nature of LLMs, repeated submissions of the same question may yield different responses, which could influence evaluation scores. Therefore, the results of this study should be interpreted as reflecting a single instance of model output rather than as a comprehensive assessment of response variability. Moreover, some of the patient questions used in this study were derived

from publicly available online sources, some of which may have been included in the model's training data. This potential overlap could have influenced the responses generated by the model. In addition, the scoring system was adapted from previously published studies and was not formally validated for the assessment of radiology-specific patient information. Furthermore, no predefined reference standard was used as the accuracy criterion, and evaluations were based on the radiologists' clinical judgment, which may introduce subjectivity. The results are specific to abdominal ultrasonography and may not be generalizable to other imaging modalities or types of ultrasonography. Despite these limitations, this study provides useful preliminary insights into the performance of LLMs in addressing common patient questions about abdominal ultrasonography.

Future studies could compare multiple LLMs to determine whether performance differs across platforms. Incorporating patient perspectives and standardized readability assessments will be important for better evaluation of response quality from the end-user perspective. Expanding the question set through broader sources such as patient forums, social media platforms, or direct clinical encounters may improve representativeness. Evaluating model performance across different languages, including English and Turkish, could also provide insight into the effect of language on response quality. In addition, larger question sets and longitudinal designs that assess consistency over time and across model updates may strengthen the evidence base. Future research may also benefit from repeated submissions of the same questions to assess response variability and improve the robustness of performance evaluation. The use of validated assessment tools, such as DISCERN or PEMAT, and the development of consensus-based reference standards for accuracy evaluation may further enhance methodological rigor. Applying similar methods to other imaging modalities, such as MRI, CT, or alternative types of ultrasonography, could help determine the generalizability of these findings across radiology subspecialties.

Conclusion

ChatGPT-5 provided responses to frequently asked patient questions about abdominal ultrasonography, which were rated favorably with consistently positive scores for relevance, accuracy, clarity, and completeness. Although the model performed well, interrater reliability showed variable levels of agreement, with moderate agreement observed for completeness and total scores. The findings indicate that LLMs may be valuable as tools to support patient education. At the same time, information generated by these systems should be viewed as complementary rather than a substitute for professional medical guidance. Integrating such tools into clinical communication may offer benefits but their use should be approached with caution. Further studies involving diverse patient groups, multiple models and broader imaging contexts will help clarify the potential role of LLMs in medical settings.

Ethics

Ethics Committee Approval: Ethical approval for this study was obtained from the İzmir Katip Çelebi University Health Research Ethics Committee (decision number: 0670, date: 06.11.2025).

Informed Consent: Due to the use of publicly available data and the absence of direct human participant involvement or identifiable patient information, the requirement for informed consent was waived by the ethics committee.

Footnotes

Authorship Contributions: Concept - A.S., G.M.T., A.D.B., R.E.B., A.H.Ç.; Design - A.S., R.E.B., A.M.K., A.H.Ç.; Data Collection or Processing - G.M.T., A.D.B., R.E.B.; Analysis or Interpretation - A.S., A.D.B., A.M.K.; Literature Search - A.S., G.M.T., A.M.K., A.H.Ç.; Writing - A.S., A.M.K., A.H.Ç.

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. Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, et al. A comprehensive overview of large language models. ACM Trans Intell Syst Technol. 2025; 16: 106:1-72.
2. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023; 29: 1930-40.
3. Artsi Y, Klang E, Collins JD, Glicksberg BS, Nadkarni GN, Korfiatis P, et al. Large language models in radiology reporting: a systematic review of performance, limitations, and clinical implications. Intell Based Med. 2025; 12: 100287.
4. Horiuchi D, Tatekawa H, Oura T, Oue S, Walston SL, Takita H, et al. Comparing the diagnostic performance of GPT-4-based ChatGPT, GPT-4V-based ChatGPT, and radiologists in challenging neuroradiology cases. Clin Neuroradiol. 2024; 34: 779-87.
5. Ayık G, Ercan N, Demirtaş Y, Yıldırım T, Çakmak G. Evaluation of ChatGPT-4o's answers to questions about hip arthroscopy from the patient perspective. Jt Dis Relat Surg. 2025; 36: 193-9.
6. Magruder ML, Rodriguez AN, Wong JCJ, Erez O, Piuzzi NS, Scuderi GR, et al. Assessing ability for ChatGPT to answer total knee arthroplasty-related questions. J Arthroplasty. 2024; 39: 2022-7.
7. Republic of Türkiye Ministry of Health GD of HIS. Health Statistics Yearbook. Available at: https://www.saglik.gov.tr/TR-84930/saglik-istatistikleri-yilliklari.html.
8. Forshaw KL, Boyes AW, Carey ML, Hall AE, Symonds M, Brown S, et al. Raised anxiety levels among outpatients preparing to undergo a medical imaging procedure: prevalence and correlates. J Am Coll Radiol. 2018; 15: 630-8.
9. Fu Y, Han P, Wang J, Shahzad F. Digital pathways to healthcare: a systematic review for unveiling the trends and insights in online health information-seeking behavior. Front Public Health. 2025; 13: 1497025.
10. Shool S, Adimi S, Saboori Amleshi R, Bitaraf E, Golpira R, Tara M. A systematic review of large language model (LLM) evaluations in clinical medicine. BMC Med Inform Decis Mak. 2025; 25: 117.
11. Einspänner E, Schwab R, Hupfeld S, Thormann M, Fuchs E, Gawlitza M, et al. Evaluating the role of large language models in supporting patient education during the informed consent process for routine radiology procedures. Br J Radiol. 2025; 98: 2184-90.
12. Khaldi A, Machayekhi S, Salvagno M, Maniaci A, Vaira LA, La Via L, et al. Accuracy of ChatGPT responses on tracheotomy for patient education. Eur Arch Otorhinolaryngol. 2024; 281: 6167-72.
13. Google Search Engine [Internet]. Mountain View (CA): Alphabet Inc.; [cited 2025 Nov 18]. Available from: https://www.google.com
14. OpenAI. ChatGPT [Internet]. San Francisco (CA): OpenAI. Available from: https://openai.com/chatgpt/

15. OpenAI. GPT-5 system card [Internet]. Available from: https://cdn.openai.com/gpt-5-system-card.pdf

16. Salbas A, Baysan EK. Assessment of large language models in musculoskeletal radiological anatomy: a comparative study with radiologists. Jt Dis Relat Surg. 2026; 37: 190-9.

17. Aydin S, Karabacak M, Vlachos V, Margetis K. Large language models in patient education: a scoping review of applications in medicine. Front Med (Lausanne). 2024; 11: 1477898.

18. Eminovic S, Levita B, Dell'Orco A, Leppig JA, Nawabi J, Penzkofer T. Comparison of multiple state-of-the-art large language models for patient education prior to CT and MRI examinations. J Pers Med. 2025; 15: 235.

19. Marey A, Saad AM, Tanas Y, Ghorab H, Niemierko J, Backer H, et al. Evaluating the accuracy and reliability of AI chatbots in patient education on cardiovascular imaging: a comparative study of ChatGPT, Gemini, and Copilot. Egypt J Radiol Nucl Med. 2025; 56: 37.

20. Gordon EB, Towbin AJ, Wingrove P, Shafique U, Haas B, Kitts AB, et al. Enhancing patient communication with Chat-GPT in radiology: evaluating the efficacy and readability of answers to common imaging-related questions. J Am Coll Radiol. 2024; 21: 353-9.

21. Kooraki S, Hosseiny M, Jalili MH, Rahsepar AA, Imanzadeh A, Kim GH, et al. Evaluation of ChatGPT-generated educational patient pamphlets for common interventional radiology procedures. Acad Radiol. 2024; 31: 4548-53.

22. Du Y, Ji C, Xu J, Wei M, Ren Y, Xia S, et al. Performance of ChatGPT and Microsoft Copilot in Bing in answering obstetric ultrasound questions and analyzing obstetric ultrasound reports. Sci Rep. 2025; 15: 14627.

23. Campbell DJ, Estephan LE, Sina EM, Mastrolonardo EV, Alapati R, Amin DR, et al. Evaluating ChatGPT responses on thyroid nodules for patient education. Thyroid. 2024; 34: 371-7.

24. Johnson D, Goodman R, Patrinely J, Stone C, Zimmerman E, Donald R, et al. Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the Chat-GPT model. Res Sq [Preprint]. 2023: rs.3.rs-2566942.

25. Yan Z, Liu J, Fan Y, Lu S, Xu D, Yang Y, et al. Ability of ChatGPT to replace doctors in patient education: cross-sectional comparative analysis of inflammatory bowel disease. J Med Internet Res. 2025; 27:e62857.

26. Gültekin O, Inoue J, Yilmaz B, Cerci MH, Kilinc BE, Yilmaz H, et al. Evaluating DeepResearch and DeepThink in anterior cruciate ligament surgery patient education: ChatGPT-4o excels in comprehensiveness, DeepSeek R1 leads in clarity and readability of orthopaedic information. Knee Surg Sports Traumatol Arthrosc. 2025; 33: 3025-31.

27. Hung YC, Chaker SC, Sigel M, Saad M, Slater ED. Comparison of patient education materials generated by chat generative pre-trained transformer versus experts: an innovative way to increase readability of patient education materials. Ann Plast Surg. 2023; 91: 409-12.

28. Hofmann HL, Vairavamurthy J. Large language model doctor: assessing the ability of ChatGPT-4 to deliver interventional radiology procedural information to patients during the consent process. CVIR Endovasc. 2024; 7: 83.

29. Mahmoudi Ghehsareh M, Asri N, Azizmohammad Looha M, Sadeghi A, Ciacci C, Rostami-Nejad M. Expert evaluation of ChatGPT accuracy and reliability for basic celiac disease frequently asked questions. Sci Rep. 2025; 15: 29871.

30. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016; 15: 155-63. Erratum in: J Chiropr Med. 2017; 16: 346.

31. Salbas A, Buyuktoka RE. Performance of large language models in recognizing brain MRI sequences: a comparative analysis of ChatGPT-4o, Claude 4 Opus, and Gemini 2.5 Pro. Diagnostics (Basel). 2025; 15: 1919.

İstanbul Med J 2026; 27(2): 149-54
DOI: 10.4274/imj.galenos.2026.03161

The Relationship Between Fibrosis-4 Score and Major Cardiovascular Events and Mortality in Patients Under 40 Years of Age with Acute Myocarditis

Emine Altuntaş, Kadir Sadıkoğlu, Kübra Balçın, Gizemnur Coşkun, Hasan Şahin, Abidin Emre Tırnaksız, Mehmet Ertürk

University of Health Sciences Türkiye, Mehmet Akif Ersoy Thoracic and Cardiovascular Surgery Training and Research Hospital, Clinic of Cardiology, İstanbul, Türkiye

ABSTRACT

Introduction: Fibrosis-4 index (Fib-4i) is a useful and practical indicator of fibrosis risk in chronic liver disease. The study aimed to explore the relationship between Fib-4i and major adverse cardiovascular events (MACEs) within one year in cases of inflammatory myopericardial syndrome (IMPS).

Methods: Between January 2018 and December 2023, 152 patients diagnosed with IMPS were stratified into two groups based on the presence of late gadolinium enhancement (LGE) on cardiac magnetic resonance imaging. The Fib-4i was calculated at admission, and MACEs due to IMPS were recorded at 12 months.

Results: Group 1 consisted of 36 LGE-negative patients, while group 2 consisted of 116 LGE-positive patients. The Fib-4i was higher in group 2 than in group 1 (p=0.014). Furthermore, it was found that aspartate aminotransferase, high-sensitivity (hs) troponin T, N-terminal pro-brain natriuretic peptide, and C-reactive protein (CRP) were higher in group 2 than in group 1 (respectively, p=0.03, p=0.002, p=0.008, p=0.014). No difference in MACEs between groups was observed at one month, six months, and one year (p=0.581, p=0.558, p=0.665, respectively). However, a positive correlation was observed between the Fib-4i and both hs-troponin T and CRP.

Conclusion: Fib-4i was higher in patients with LGE than in those without LGE. However, its predictive power for MACEs could not be demonstrated.

Keywords: Fibrosis-4 index, inflammatory myopericardial syndrome, cardiac magnetic resonance, major cardiovascular events

Introduction

Inflammatory myopericardial syndrome (IMPS) is a newly described syndrome. This encompasses myocarditis and pericarditis. This description was first defined in the 2025 European Society of Cardiology (ESC) guideline on myocarditis and pericarditis. The clinical spectrum of IMPS encompasses isolated myocarditis and pericarditis, as well as combined forms, including myopericarditis and perimyocarditis. Early-onset coronary artery disease manifests in individuals under 40 and can be diagnostically challenging, particularly when it presents with symptoms similar to those of myocarditis. The aetiology of this clinical scenario is multifaceted, with potential causes including exposure to toxic substances, infectious agents, or inflammatory conditions. This diagnosis indicates a serious, potentially underdiagnosed condition that affects individuals of all ages. The clinical manifestations are heterogeneous and could overlap significantly with other acute cardiac conditions. That causes complications in the diagnostic process (1-3).

In previous consensus statements, the diagnosis of myocarditis was based on the finding on endomyocardial biopsy (EMB). EMB is an effective diagnostic tool for identifying the histological type of the disease, determining specific aetiologies, and differentiating between IMPS and other conditions. However, multimodality imaging has become a cornerstone for diagnosis, with cardiovascular magnetic resonance imaging (CMRI) playing a crucial role (4,5).

Fibrosis-4 index (Fib-4i), a non-invasive marker of hepatic fibrosis in patients with a viral infection, is calculated using three biochemical values and age (6,7). In recent years, the validity of this score has been demonstrated in liver diseases. However, other studies have

Address for Correspondence: Assoc. Prof. Emine Altuntaş, MD, University of Health Sciences Türkiye, Mehmet Akif Ersoy Thoracic and Cardiovascular Surgery Training and Research Hospital, Clinic of Cardiology, İstanbul, Türkiye
E-mail: emine_altuntas@hotmail.com **ORCID ID:** orcid.org/0000-0001-5887-5422

Cite this article as: Altuntaş E, Sadıkoğlu K, Balçın K, Coşkun G, Şahin H, Tırnaksız AE, et al. The relationship between Fibrosis-4 score and major cardiovascular events and mortality in patients under 40 years of age with acute myocarditis. İstanbul Med J. 2026; 27(2): 149-54

Received: 08.12.2025
Accepted: 17.04.2026
Publication Date: 12.05.2026

suggested that, in patients with non-hepatic primary diseases, this index is significantly associated with poor clinical outcomes. It has been demonstrated by several recent studies that Fib-4i is a valuable predictor of prognosis in patients suffering from coronary artery disease, heart failure (HF), or atrial fibrillation (6,8,9). Nevertheless, correlation between Fib-4i, and its clinical implications in IMPS patients remains to be elucidated.

The aim of this study was to ascertain whether the Fib-4i increased in IMPS patients undergoing CMRI and to determine whether there was a relationship between the Fib-4i score and prognosis in the clinical follow-up of these patients.

Methods

Study Subjects and Protocol

The study was designed as a retrospective cross-sectional investigation. Between January 2018 and December 2023, patients hospitalized with a provisional diagnosis of acute coronary syndrome who were under the age of 40 were selected. Coronary angiography or coronary computed tomography was performed to exclude coronary artery disease in these patients. Patients who satisfied the diagnostic criteria outlined in 2025 ESC guidelines for myocarditis and pericarditis were evaluated as having IMPS (2). Patients who had undergone CMRI were included in the study. During follow-up, death, rehospitalization due to HF or arrhythmia, and recurrent IMPS were recorded at the 1-month, 6-month, and 1-year outpatient clinic visits. The major cardiovascular events (MACEs) in this study constituted a composite endpoint. Participants were excluded from the study if they were under the age of 18, pregnant, had a history of acute or chronic hepatitis B or C, or had liver cirrhosis due to any cause. Also, patients who did not complete follow-up or who had missing data were excluded from the study. Medical records indicated that all patients had experienced an acute infection before the establishment of the IMPS clinic. The present study comprised 152 patients. Patients were divided into two groups based on CMRI involvement: group 1, without CMRI involvement (n=36), and group 2, with CMRI involvement (n=116). Patient data were retrieved from hospital medical records.

All patients were treated from admission with ibuprofen (600-800 mg, 3 x 1), proton pump inhibitors (PPIs), and colchicine (0.5 mg, 2 x 1). Ibuprofen and PPI treatments were given for 2 weeks, while colchicine treatment was given for at least 3 months. For patients with recurrent IMPS, the duration of colchicine treatment has been extended to a maximum of six months.

The study protocol adhered to the ethical guidelines of the Declaration of Helsinki and was approved by the Ethics Committee of the University of Health Sciences Türkiye, Mehmet Akif Ersoy Thoracic and Cardiovascular Surgery Training and Research Hospital (approval number: 2025.01-04, date: 14.01.2025). Because of the retrospective design of the study, individual informed consent was waived.

Laboratory Test and Fib-4i Calculation

On admission, venous blood samples were drawn from each patient and analyzed. N-terminal pro B-type natriuretic peptide (NT-proBNP) and high-sensitivity (hs) troponin-T were measured on admission. An automated biochemical analyzer (Roche COBAS 6000) was used to measure the concentrations of total cholesterol (TC), fasting blood glucose (FBG), aspartate aminotransferase (AST), alanine aminotransferase (ALT), and serum creatinine, using an enzymatic assay. A complete blood count was performed using the Mindray BC-6000 autoanalyzer.

The Fib-4i at admission was calculated by the formula: age (years) × AST(U/L)/(platelets count [10^9/μL] × $\sqrt{\text{ALT}}$ (U/L) (6).

Trans Thoracic Echocardiography

Echocardiographic assessments were performed at baseline by two independent, experienced physicians who were blinded to the results. All parameters were assessed following the current guidelines of the American Society of Echocardiography and the ESC (10,11).

CMRI Acquisition Protocol

CMRI was performed using a 1.5-Tesla MR scanner (Magnetom Aera; Siemens Healthcare, Erlangen, Germany). Cine imaging was obtained with a breath-hold balanced steady-state free precession sequence in standard long-axis planes (two-, three-, and four-chamber views) and contiguous short-axis orientations (slice thickness of 6–8 mm, without interslice gap; 10–15 slices). Myocardial edema was assessed using T2-weighted short tau inversion recovery sequences in combination with T2 mapping.

Native T1 mapping was acquired using an electrocardiography-triggered modified Look-Locker inversion recovery sequence with a 3(3)3(3)5 scheme at three short-axis levels of the left ventricle (basal, mid-ventricular, and apical). After intravenous injection of gadobutrol at a dose of 0.15 mmol/kg, late gadolinium enhancement (LGE) images were obtained using a two-dimensional breath-hold phase-sensitive inversion-recovery gradient-echo sequence in matching imaging planes. The inversion time was individually adjusted to suppress the signal from normal myocardium. Post-contrast T1 mapping was performed following LGE imaging.

Statistical Analysis

Statistical analysis was implemented using SPSS 22.0 (USA, Amonk, NY, IBM Corporation) from gathered data analysis, continuous factors (CFs) are stated as mean ± standard deviation, and categorical factors are stated as a percentage of group total %. The Kolmogorov–Smirnov test was utilized to determine whether the factors demonstrated a normal distribution. CFs with a normal distribution were evaluated using Student's t-test. The chi-square test was used for categorical variables. Correlation analyses were performed using Pearson or Spearman correlation coefficients as appropriate. A p value <0.05 was considered significant.

Results

Patients were divided into two groups based on the presence of LGE on CMRI: group 1 (n=36) included patients without LGE, and group 2 (n=116) included patients with LGE. Group 1 had a mean age of 24.5 years and consisted of 24 men (66.7%), while group 2 had a mean age of 21 years and consisted of 89 men (76.7%). The groups were similar in terms of age and gender (p=0.051 and p=0.161, respectively). The patients

had no chronic illnesses. One patient in group 1 and eight patients in group 2 were smokers (p=0.325). Furthermore, groups were compared with respect to symptoms presenting to the emergency department; no statistically significant difference was detected (p>0.05). The groups were also compared with respect to laboratory tests. AST, C-reactive protein (CRP), hs-troponin T, NT-proBNP, and Fib-4i were higher in group 2 than in group 1, and these differences were statistically significant (p=0.03, p=0.044, p=0.002, p=0.008, p=0.014, respectively). The laboratory test results were summarised in Table 1.

The left ventricular ejection fraction (LVEF) in patients was assessed using both transthoracic echocardiography and CMRI, and the results were compared. The groups showed similar LVEF levels (p=0.771 and p=0.851). Furthermore, a comparative analysis of MACE between the two groups was conducted. Recurrent episodes of myocarditis were observed as MACE in the groups. No further events were observed. Again, no statistically significant difference was observed between the groups with respect to MACE at 1, 6, and 12 months (p=0.581, p=0.558, and p=0.665, respectively). The results are summarized in Table 2.

A correlation analysis was performed among the Fib-4i, NT-proBNP, hs-troponin T, TC, and FBG. The present study found a positive correlation between the FIB-4 index and hs-troponin T and CRP. The results of the correlation analysis were presented in Table 3, Figures 1 and 2.

Table 1. Comparison between two groups in terms of demographical, clinical, and laboratory tests results

Variables	Total (n=152)	Group 1 (n=36)	Group 2 (n=116)	p
Gender (male, %)	113 (74.3%)	24 (66.7%)	89 (76.7%)	0.161*
Age (years)	31 (25.5-34.5)	24.5 (19.7-33.25)	21 (19-29)	0.051**
Smoking	9 (5.9%)	1 (2.8%)	8 (6.9%)	0.325*
Admission complaint				
Chest pain	143 (93.6%)	32 (88.9%)	111 (94.8%)	0.207*
Dyspnea	4 (2.6%)	2 (5.6%)	2 (1.7%)	
Syncope	3 (2%)	1 (2.8%)	2 (1.7%)	
Palpitation	2 (1.3%)	1 (2.8%)	1 (0.9%)	
Creatinine (mg/dL)	0.73 (0.66-0.81)	0.8 (0.74-0.9)	0.79 (0.63-0.86)	0.082**
Hb (g/dL)	11.9 (11.0-12.9)	14.65(12.8-15.4)	14.2 (13.3-15)	0.713**
WBC (10^9/L)	10 (9.73-10.85)	9.3 (7-10.8)	8.13 (6.5-10)	0.319**
PLT (10^9/L)	214 (195-217)	238.5 (210.25-267)	218 (196.75-266)	0.152**
AST (U/L)	33 (23-60)	23.5 (17-38.5)	39 (25-63)	0.03**
ALT (U/L)	19 (13-27)	19,5 (12.25-28.75)	19 (13-27)	0.804**
Fasting glucose (mg/dL)	100.5 (89-107)	98.5 (85-110)	101 (89-107)	0.822**
Total cholesterol (mg/dL)	150 (141-150.5)	154 (97-185)	138 (113-152)	0.322**
hs-Troponin T (ng/L)	147.2 (92.1-673.6)	109 (29.25-384.75)	408 (111.25-860)	0.002**
CRP (mg/L)	10 (5.1-20.5)	5.5 (2-38.6)	28.12 (8-46)	0.044**
NT-BNP (pg/mL)	193 (103.5-320)	122.5 (97.5-233.5)	207 (109-361)	0.008**
Fib-4 index	0.9 (0.5-1.26)	0.53 (0.4-0.96)	0.97 (0.62-1.34)	0.014**

*: Chi-square test, **: Mann-Whitney U test, Hb: Hemoglobin, WBC: White blood cell, PLT: Platelet, AST: Aspartate aminotransferase, ALT: Alanine aminotransferase, hs-troponin T: High sensitivity Troponin T, CRP: C-reactive protein, NT-BNP: N-terminal pro b-type natriuretic peptide, Fib-4 index: Fibrosis-4 index, dL: Deciliter, mg: Milligram, mL: Milliliter, L: Liter, ng: Nanogram, pg: Picogram

Table 2. Comparison between two groups in terms of trans thoracic echocardiographic, cardiac magnetic resonance imaging test results, and major cardiac events

Variables	Total (n=152)	Group 1 (n=36)	Group 2 (n=116)	p
EF on admission (%)	60 (60-60)	60 (58.7-60)	60 (60-60)	0.771*
EF measured by cardiac MRI (%)	61 (59-64)	60 (60-63)	61 (59-65)	0.865*
First month MACE	2 (1.3%)	0	2 (1.7%)	0.581**
Sixth month MACE	3 (2%)	1 (2.8%)	2 (1.77%)	0.558**
First year MACE	4 (2.6%)	1 (2.8%)	3 (2.6%)	0.665**

*: Mann-Whitney U test, **: Chi-square test. EF: Ejection fraction, MACE: Major cardiovascular events, mortality, heart failure, rehospitalization due to decompensated heart failure, and arrhythmia, recurrent myocarditis attack, MRI: Magnetic resonance imaging

Table 3. Spearman's correlation analysis among Fib-4 index and some laboratory tests results

Variables	Fib-4 index	NT-pro BNP	hs-troponin T	Total cholesterol	Fasting glucose level	CRP
Fib-4	1	p= 0.843 r= 0.020	p= <0.001 r= 0.683	p= 0.147 r= 0.195	p= 0.090 r= 0.176	p= <0.001 r= 0.397
NT-proBNP (pg/mL)		1	p= 0.854 r= 0.017	p= 0.067 r= 0.238	p= 0.489 r= 0.070	p= 0.798 r= 0.025
hs-troponin T (ng/L)			1	p= 0.483 r= 0.093	p= 0.073 r= 0.181	p= <0.001 r= 0.464
Total cholesterol (mg/dL)				1	p= 0.532 r= 0.082	p= 0.250 r= 0.155
Fasting glucose level (mg/dL)					1	p= 0.759 r= 0.032
CRP (mg/L)						1

Fib-4: Fibrosis-4, NT-proBNP: N-terminal pro b-type natriuretic peptide, hs-troponin T: High sensitivity troponin T, CRP: C- reactive protein, dL: deciliter, L: liter, mg: Miligram, mL: Milliliter, ng: Nanograms, pg: picogram, r: Correlation coefficient

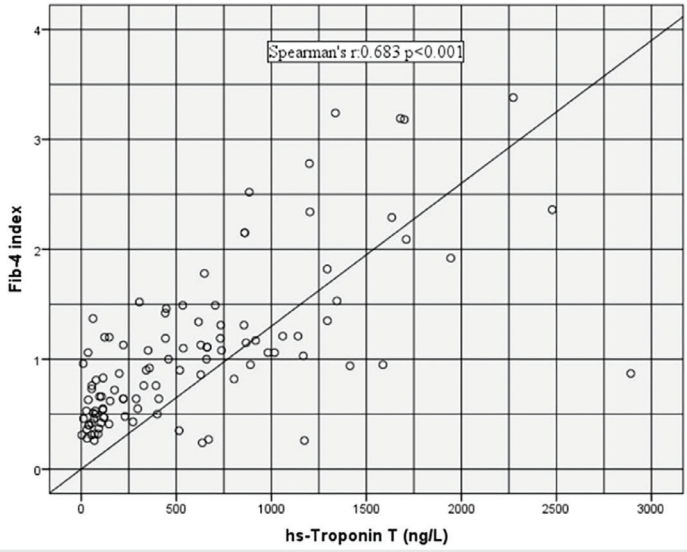

Figure 1. Spearman's correlation analysis between Fib-4 index and hs-troponin T. Fib-4: Fibrosis-4, hs-troponin T: High sensitivity troponin T

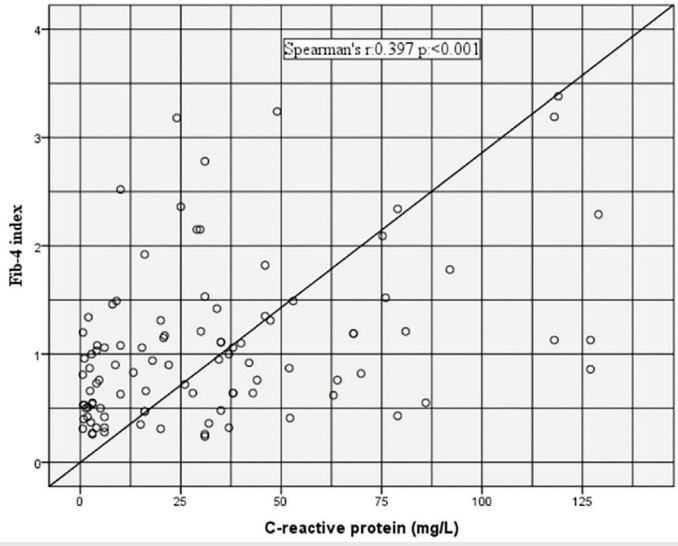

Figure 2. Spearman's correlation analysis between Fib-4 index and CRP. Fib-4: Fibrosis-4, CRP: C- reactive protein

Discussion

This study investigated the relationship between Fib-4i and 1-year MACEs in IMPS patients. Fib-4i was higher in LGE-positive patients. Despite the absence of a demonstrable relationship between this elevation and MACEs, a positive correlation was identified between the Fib-4i and hs-troponin T.

Fib-4i has been demonstrated to be a significant predictor in various clinical scenarios. The increased index is an independent risk factor for diabetic neuropathy and chronic kidney disease in patients with type 2 diabetes, hepatocellular carcinoma in patients with chronic hepatitis B, in patients with atrial fibrillation-related stroke, and hospitalized acute HF (12-16). Despite the general prognostic ability of Fib-4i, it did not emerge as an independent prognostic factor in outcome of this study. The absence of HF, arrhythmia, and pericardial effusion in patients may have led to this outcome. In a multicenter study involving 1,162 patients hospitalized due to acute HF, it was demonstrated that Fib-4i calculated at admission was an important predictor of both all-cause mortality, and rehospitalization (16).

In another study, 704 HF patients were followed for 5 years. The patients were divided into three groups: HF with preserved EF, HF with mildly reduced EF, and HF with reduced EF. The primary outcome was a composite of total cardiovascular events (CVEs). Cardiovascular-related death; hospitalization for HF decompensation; non-fatal MI; unstable angina pectoris; coronary revascularization for a new diagnosis of angina or for in-stent restenosis after percutaneous coronary intervention; and non-fatal ischemic stroke were defined as total CVEs. In study, it was found that, Fib-4i was a substantial predictor for total CVEs in HF population (6). In both patient groups, systemic venous congestion increases neurohormonal activation. This activation leads to HF progression and may contribute to more severe multiple organ failure, resulting in a poor prognosis. HF can involve congestion and reduced arterial flow, resulting in hypoxic hepatopathy. Hypoxia can cause centrilobular necrosis of the liver, leading to elevated transaminase levels. Increased central venous pressure also causes hepatocyte atrophy and perisinusoidal oedema. Evidence suggests a link between liver congestion and liver stiffness. This can lead to fibrosis. This can result in a poor prognosis (6,17-19). IMPS represents a clinical spectrum. It can range from a silent clinical course to fulminant hepatitis or constrictive pericarditis. If the clinical presentation leads to non-ischaemic dilated HF or right HF secondary to constrictive pericarditis, fibrosis is triggered by the aforementioned pathophysiology. However, hepatomegaly, splenomegaly, or both may be present in patients, and reduced platelet counts may also occur. Consequently, the FIB-4 index may be elevated in these patients (2,4,6,16-18).

Our study established positive correlations between Fib-4i and hs-troponin T and between Fib-4i and CRP. In a retrospective analysis, the CMRI data and laboratory parameters of 244 patients with clinical suspicion of acute myocarditis were assessed. Analysis demonstrated that hs-troponin T levels ≤18 pg/mL corresponded to a very low-risk of acute myocarditis, suggesting that CMRI may not be necessary for exclusion (19). Another study assessed the correlation between hs-troponin I levels and myocardial damage on CMRI, represented by LGE percentage, in patients diagnosed with myocarditis. The study consisted of 101 patients. They found a linear association between LGE percentage and maximal hs troponin I value with r: 0.49 ($p<0.001$) (20). In light of this information, two hypotheses can be established: 1) In patients with negative hs-troponin T values and no other clinical conditions, the Fib-4i may also remain low. However, a cut-off value must be determined for this purpose. 2) A positive correlation between hs-troponin T and the Fib-4i may also exist between LGE percentage and Fib-4i. However, additional studies are required to substantiate these findings.

Study Limitations

The study had several limitations: 1) Sample size was relatively small. A more extensive study is required to determine whether the Fib-4i is an effective predictor of MACE. 2) The study was conducted at a single center and was observational. The inclusion of a single population and a single center in the study limits the generalisability of the results and suggests that further research is required to assess predictive value of Fib-4i in other populations. 3) Sample population comprised young patients with uncomplicated IMPS. A more homogeneous population was required to enable a more accurate assessment. 4) Another limitation is the lack of quantitative assessment of the LGE ratio on CMRI.

Conclusion

This study demonstrated that the FIB-4 index was higher in IMPS patients with LGE detected on CMRI, and there was a positive correlation between troponin and the FIB-4 index. Therefore, Fib-4i could be useful for assessing clinical risk in this patient group.

Ethics

Ethics Committee Approval: The study was approved by the Ethics Committee of the University of Health Sciences Türkiye, Mehmet Akif Ersoy Thoracic and Cardiovascular Surgery Training and Research Hospital (approval number: 2025.01-04, date: 14.01.2025).

Informed Consent: Because of the retrospective design of the study, individual informed consent was waived.

Footnotes

Authorship Contributions: Surgical and Medical Practices - E.A., K.S., K.B., G.C., H.Ş., A.E.T.; Concept - E.A., M.E.; Design - E.A., M.E.; Data Collection or Processing - K.S., K.B., G.C., H.Ş., A.E.T.; Analysis or Interpretation - E.A., M.E.; Literature Search - E.A., K.S., K.B., G.C., H.Ş., A.E.T.; Writing - E.A.

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. Lampejo T, Durkin SM, Bhatt N, Guttmann O. Acute myocarditis: aetiology, diagnosis and management. Clin Med (Lond). 2021; 21: e505-10.
2. Schulz-Menger J, Collini V, Gröschel J, Adler Y, Brucato A, Christian V, et al. 2025 ESC Guidelines for the management of myocarditis and pericarditis. Eur Heart J. 2025; 46: 3952-4041.

3. Usalp S, Altuntaş E, Bağırtan B, Karabay KÖ. Comparison of serum lipoprotein(a) levels in young and middle-aged patients presenting for the first time with ST-elevation myocardial infarction: a single-centre study. Cardiovasc J Afr. 2023; 34: 1-5.

4. Writing Committee; Drazner MH, Bozkurt B, Cooper LT, Aggarwal NR, Basso C, Bhave NM, et al. 2024 ACC Expert Consensus decision pathway on strategies and criteria for the diagnosis and management of myocarditis: a report of the American College of Cardiology Solution Set Oversight Committee. J Am Coll Cardiol. 2025; 85: 391-431.

5. Caforio AL, Pankuweit S, Arbustini E, Basso C, Gimeno-Blanes J, Felix SB, et al. Current state of knowledge on aetiology, diagnosis, management, and therapy of myocarditis: a position statement of the European Society of Cardiology Working Group on Myocardial and Pericardial Diseases. Eur Heart J. 2013; 34: 2636-48, 2648a-2648d.

6. Takae M, Fujisue K, Yamamoto E, Egashira K, Komorita T, Oike F, et al. Prognostic significance of liver stiffness assessed by fibrosis-4 index in patients with heart failure. ESC Heart Fail. 2021; 8: 3809-21.

7. Raposeiras-Roubín S, Parada Barcia JA, Lizancos Castro A, Noriega Caro V, Ledo Piñeiro A, González Bermúdez I, et al. Liver fibrosis and outcomes of atrial fibrillation: the FIB-4 index. Clin Res Cardiol. 2024; 113: 313-23.

8. McNally BB, Rangan P, Wijarnpreecha K, Fallon MB. Fibrosis-4 index score predicts concomitant coronary artery diseases across the spectrum of fatty liver disease. Dig Dis Sci. 2023; 68: 3765-73.

9. Sudo M, Shamekhi J, Sedaghat A, Aksoy A, Zietzer A, Tanaka T, et al. Predictive value of the Fibrosis-4 index in patients with severe aortic stenosis undergoing transcatheter aortic valve replacement. Clin Res Cardiol. 2022; 111: 1367-76.

10. Wiegers SE, Ryan T, Arrighi JA, Brown SM, Canaday B, Damp JB, et al. 2019 ACC/AHA/ASE advanced training statement on echocardiography (Revision of the 2003 ACC/AHA clinical competence statement on echocardiography): a report of the ACC Competency Management Committee. J Am Coll Cardiol. 2019; 74: 377-402.

11. Galderisi M, Cosyns B, Edvardsen T, Cardim N, Delgado V, Di Salvo G, et al. Standardization of adult transthoracic echocardiography reporting in agreement with recent chamber quantification, diastolic function, and heart valve disease recommendations: an expert consensus document of the European Association of Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging. 2017; 18: 1301-10.

12. Kim K, Oh TJ, Cho HC, Lee YK, Ahn CH, Koo BK, et al. Liver fibrosis indices are related to diabetic peripheral neuropathy in individuals with type 2 diabetes. Sci Rep. 2021; 11: 24372.

13. Seko Y, Yano K, Takahashi A, Okishio S, Kataoka S, Okuda K, et al. FIB-4 index and diabetes mellitus are associated with chronic kidney disease in Japanese patients with non-alcoholic fatty liver disease. Int J Mol Sci. 2019; 21: 171.

14. Suzuki T, Matsuura K, Nagura Y, Iio E, Ogawa S, Fujiwara K, et al. Development of hepatocellular carcinoma from various phases of chronic hepatitis B virus infection. PLoS One. 2021; 16: e0261878.

15. Kim TH, Kim SY, Jung YK, Yim HJ, Jung JM, Seo WK. FIB-4 index and liver fibrosis are risk factors for long-term outcomes in atrial fibrillation-related stroke. Clin Neurol Neurosurg. 2022; 217: 107235.

16. Shibata N, Kondo T, Kazama S, Kimura Y, Oishi H, Arao Y, et al. Impact of predictive value of Fibrosis-4 index in patients hospitalized for acute heart failure. Int J Cardiol. 2021; 324: 90-95.

17. Jalal Z, Iriart X, De Lédinghen V, Barnetche T, Hiriart JB, Vergniol J, et al. Liver stiffness measurements for evaluation of central venous pressure in congenital heart diseases. Heart. 2015; 101: 1499-504.

18. Mohamed BA, Schnelle M, Khadjeh S, Lbik D, Herwig M, Linke WA, et al. Molecular and structural transition mechanisms in long-term volume overload. Eur J Heart Fail. 2016; 18: 362-71.

19. Puetz A, Berger M, Lebherz C, Bauermann K, Hartmann NU, Kappel BA, et al. High-sensitivity troponin T as a rule-out marker for myocardial inflammation detectable by CMR imaging. Open Heart. 2025; 12: e003508.

20. Marcusohn E, Barbara A, Epstein D, Massalha S, Zukermann R. Correlations between high sensitive troponin I and acute myocarditis extent in cardiac magnetic resonance imaging. J Cardiovasc Med (Hagerstown). 2023; 24: 334-9.

İstanbul Med J 2026; 27(2): 155-8
DOI: 10.4274/imj.galenos.2026.65288

Cytologic Evaluation of Papillary Breast Lesions: A 35-Case Series with Histopathologic Correlation and Clinical Follow-Up

Burcu Özcan, Ayşenur Başkan

University of Health Sciences Türkiye, İstanbul Training and Research Hospital, Clinic of Pathology, İstanbul, Türkiye

ABSTRACT

Introduction: Papillary lesions of the breast constitute a heterogeneous group ranging from intraductal papillomas to papillary carcinomas. Fine-needle aspiration cytology (FNAC) plays an important role in the initial evaluation of these lesions; however, cytologic distinction between benign and malignant papillary lesions is challenging due to overlapping morphologic features. The aim of this study was to determine the diagnostic value of cytologic findings and the associated risk of neoplasia (RON) and malignancy in papillary breast lesions.

Methods: A total of 35 cases of breast FNAC and nipple discharge cytology that demonstrated papillary architecture (2017–2024) were retrospectively evaluated. Cytomorphologic assessment focused on two diagnostic features—nuclear atypia and hypercellularity—and their correlation with histologic outcomes. The RON and the risk of malignancy (ROM) were calculated with 95% confidence intervals (CIs). Statistical analysis was performed using Fisher's exact test.

Results: All patients were female (age range: 25–71 years; mean: 47 years). Histopathologic diagnosis was available for 8 cases, and clinical follow-up for 17 cases. Nuclear atypia was identified in 5 cases (14%), while hypercellularity was observed in 17 cases (48.5%). RON and ROM were 60% and 20% in cases with nuclear atypia and 35.3% and 11.8% in cases with hypercellularity, respectively. Among all evaluable cases, overall RON was 32.0% (95% CI: 17.7–51.6) and ROM was 12.0% (95% CI: 4.2–30.0).

Conclusion: FNAC remains a valuable and minimally invasive diagnostic tool in the evaluation of papillary breast lesions. Although definitive cytologic distinction between benign and malignant lesions is not always possible, the identification of nuclear atypia and hypercellularity serves as an important indicator of neoplastic potential. The integration of immunocytochemical studies and standardized reporting systems may further improve diagnostic accuracy and reproducibility in the cytological evaluation of papillary breast lesions.

Keywords: Papillary breast lesion, fine-needle aspiration cytology, nuclear atypia, hypercellularity, risk of neoplasia, risk of malignancy

Introduction

Papillary lesions of the breast encompass a varied spectrum of neoplasms characterized by the presence of fibrovascular cores lined by epithelial cells, forming intricate papillary architectures within the ductal-lobular system. These lesions range from benign entities such as intraductal papillomas to malignant forms including encapsulated papillary carcinoma and invasive papillary carcinomas (1-3). The morphological heterogeneity inherent in this group poses significant diagnostic challenges, especially when differentiating benign from malignant lesions on cytology. Indeed, the overlapping cytomorphologic features between benign papillomas and papillary carcinomas often lead to difficulties in definitive preoperative categorization, which is critical for guiding patient management decisions (1,4).

Epidemiologically, breast papillary lesions occur predominantly in middle-aged and elderly women, although rare cases in men have been documented, further broadening the clinical presentation and diagnostic considerations. Clinically, such lesions frequently present as palpable masses or with nipple discharge; both findings prompt imaging and tissue sampling. The accurate diagnosis of papillary lesions is essential since, despite many being indolent, some can harbor areas of atypia, carcinoma in situ, or invasive carcinoma components that significantly influence prognosis and therapeutic strategy (4).

Fine needle aspiration cytology (FNAC) represents a minimally invasive, cost-effective, and rapid method for the initial evaluation of palpable and image-identified breast lesions. In resource-limited or busy clinical settings, FNAC remains a valuable diagnostic tool for breast lesions,

Address for Correspondence: Burcu Özcan, MD, University of Health Sciences Türkiye, İstanbul Training and Research Hospital, Clinic of Pathology, İstanbul, Türkiye
E-mail: drburcuozcan@yahoo.com **ORCID ID:** orcid.org/0000-0002-7662-3306

Cite this article as: Özcan B, Başkan A. Cytologic evaluation of papillary breast lesions: a 35-case series with histopathologic correlation and clinical follow-up. İstanbul Med J. 2026; 27(2): 155-8

Received: 17.03.2026
Accepted: 26.04.2026
Publication Date: 12.05.2026

including papillary neoplasms, because of its simplicity and low complication rates. However, for papillary lesions, despite FNAC's established role, the technique is subject to inherent limitations related to sampling and interpretative challenges due to the complex architecture and overlapping cytological features found in these lesions (5).

Several studies underline the pitfalls encountered with FNAC in diagnosing papillary lesions. False positives often arise from overinterpretation of reactive or atypical changes or misidentification of papillaroid fragments in non-papillary lesions such as fibroadenomas or certain invasive ductal carcinomas that present with papillary-like morphology. Conversely, false negatives may occur due to hypocellularity, scant fibrovascular cores, or sampling errors, especially when true papillary structures are absent or poorly represented in aspirates (4).

The sensitivity of FNAC in recognizing papillary breast lesions is variable; some series report figures near 54%, which reflects the cautious approach required when interpreting aspirates for such lesions. The identification and careful evaluation of true papillae, characterized by epithelial clusters surrounding fibrovascular cores, columnar cell morphology, and cellular atypia, remain cornerstones for cytologic diagnosis. Experts recommend that the presence of complex branching papillary fragments with fibrovascular cores and assessment of nuclear features such as mild to severe atypia might aid in distinguishing benign from malignant papillary lesions (4).

The present study aimed to evaluate the predictive value of nuclear atypia and hypercellularity in cytologically diagnosed papillary breast lesions and to determine the corresponding risks of neoplasia (RON) and risk of malignancy (ROM).

Methods

This retrospective single-center study included cases of breast FNAC and nipple discharge cytology that were evaluated in our pathology department between August 2017 and September 2024. Among these cases, those demonstrating papillary structures on cytologic examination were selected (n=35). Demographic data on the patients were retrieved from the hospital information system. Cases in which the cytologic material lacked papillary architecture or was deemed inadequate were excluded from the study. All cases were reviewed by a single experienced cytopathologist to ensure diagnostic consistency.

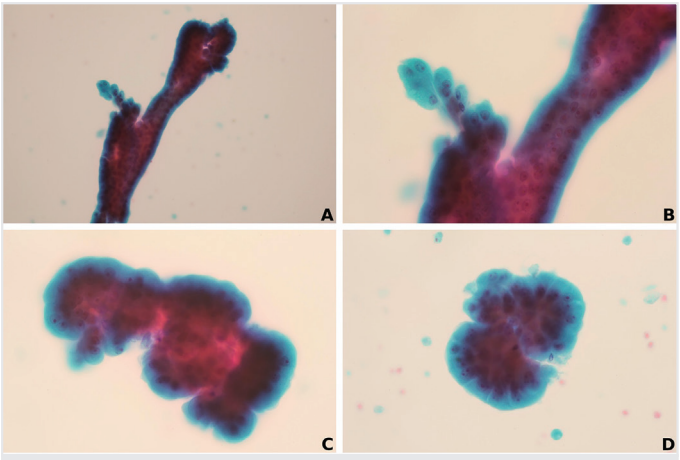

Figure 1. A) A well-formed papillary structure with ductal epithelial cells arranged in a single row around a fibrovascular core (PAP, × 400). B) High-magnification view of the cells forming the papillary structure. No nuclear atypia is identified (smooth nuclear contours and fine chromatin). (PAP ×1000). C) Complex three-dimensional crowded cell clusters detached from the tips of papillary structures (PAP, × 1000). D) A rounded cell cluster with contours resembling a papillary cap (PAP, × 1000)
PAP: Papanicolaou

All specimens were submitted as fluid aspirates and processed using a liquid-based cytology (LBC) technique. The primary LBC slides were prepared using the SurePath Pap Test kit (BD Diagnostics). Specimens were fixed in an ethanol-based solution (CytoRich™ Red, BD Diagnostics) and subjected to two centrifugation steps. The cellular material was subsequently vortexed to ensure homogenization, and then was evenly distributed as a thin layer on microscope slides. A single smear was prepared from each case and stained with the Papanicolaou method. Cell block material was available for all cases.

All cases demonstrated papillary structures on cytologic evaluation (Figure 1). All cases were further assessed for two key diagnostic features: (1) nuclear atypia, defined by the presence of nuclear enlargement, irregular nuclear contours, and coarse chromatin, either individually or in combination; and (2) hypercellularity, reflecting the proliferative nature of the lesion.

The impact of these cytologic features on the RON and ROM. was analyzed. For each case, the availability of histopathologic diagnosis was recorded, and cytologic–histologic concordance was assessed. Immunohistochemical evaluation was performed only on available histologic sections from selected cases and was not applied to cytologic material.

Ethical approval for the study was granted by the University of Health Sciences Türkiye, İstanbul Training and Research Hospital Ethics Committee (decision number: 303, date: 05.12.2025). Due to the retrospective nature of the study, the requirement for informed consent was waived by the ethics committee.

Statistical Analysis

RON and ROM values were calculated with 95% Wilson confidence intervals (CIs). The association between nuclear atypia and malignancy was analyzed using Fisher's exact test due to the small sample size. Statistical analyses were performed using SPSS version 26.0 (IBM Corp., Armonk, NY, USA). A p value <0.05 was considered statistically significant.

Results

The patients ranged in age from 25 to 71 years (median: 45 years; mean: 47 years), and all were female. The lesions were located in the right breast in 12 cases (34%) and in the left breast in 23 cases (66%). Periareolar localization was observed in 14 cases. Lesion size varied between 5 and 60 mm (mean: 22 mm). Thirty-one cases were obtained by fine-needle aspiration, while four were smear preparations of nipple discharge. All cases were cytologically diagnosed as intraductal papillary lesions, and the presence or absence of atypical features was documented in the reports.

Histopathologic diagnosis was available for eight cases (seven excisional specimens and one tru-cut biopsy). Cytomorphologic evaluation revealed nuclear atypia in 5 of 35 cases (14%). Among these, three had histologic follow-up: two were diagnosed as intraductal papilloma, and one as ductal carcinoma in situ with a papillary pattern. A closer examination of the cytologic features of these three cases revealed that nuclear atypia was focal and mild in the two cases with histologic diagnoses of intraductal papilloma. In contrast, the case diagnosed as ductal carcinoma in situ exhibited mild-to-moderate nuclear atypia. Two additional cases were followed clinically for 9 months and 22 months, respectively, without recurrence.

RON among cases with nuclear atypia was 60%, and the ROM was 20%. The distribution of cases by cytologic features and associated risk rates is summarized in Table 1. There was no statistically significant difference in malignancy frequency between cases with atypia and those lacking atypia (Fisher's exact test, p=0.544).

Hypercellularity, reflecting proliferative activity, was identified in 17/35 cases (48.5%). Six of these cases underwent histologic evaluation, revealing four intraductal papillomas, one invasive ductal carcinoma, and one ductal carcinoma in situ. The remaining 10 cases were followed clinically for 2–56 months (mean, 32 months), and no recurrence or suspicious radiologic findings were detected. The calculated RON and ROM for proliferative cases were 35.3% and 11.8%, respectively. Detailed correlations between cytologic findings and histologic or clinical follow-up are presented in Table 2. The association between proliferation and malignancy was not statistically significant (Fisher's exact test, p=1.000).

Both cytologic features—nuclear atypia and hypercellularity—were concurrently observed in three cases (8.5%). Two of these were diagnosed as intraductal papillomas on excision, and one case showed no recurrence during a 9-month follow-up. All cases showing both features were histologically neoplastic (RON 100%), as summarized in Table 1.

Ten cases were lost to follow-up or had follow-up periods shorter than six months. Seventeen cases were clinically monitored for 9–73 months (mean, 32 months), and none developed suspicious lesions or malignancy during the follow-up period.

Overall, cytologic evaluation identified 35 cases consistent with intraductal papillary lesions. Of these, 25 had either histopathologic correlation (n=8) or sufficient clinical follow-up (n=17). All tissue samples demonstrated neoplastic lesions, including three malignant cases. Based on evaluable cases, the RON was 32.0% (95% CI: 17.7–51.6) and the ROM was 12.0% (95% CI: 4.2–30.0) (Table 3). When all cytologically diagnosed cases were considered, the overall RON and ROM were 22.9% and 8.6%, respectively.

Immunohistochemical studies were not performed on cytologic material; however, in five cases of intraductal papilloma, myoepithelial markers (p63, calponin, and CK5/6) were applied to the corresponding tissue sections, confirming an intact myoepithelial cell layer.

Discussion

The cytologic diagnosis of papillary breast lesions remains one of the most challenging aspects of breast cytopathology because of the substantial morphologic overlap between benign and malignant entities. All histologically correlated cases in our series were found to be neoplastic. The calculated RON was 32%, and the ROM was 12%, values consistent with previously reported malignancy rates ranging between 8% and 20% in similar series (4,6).

The clinical management of papillary breast lesions depends heavily on accurate diagnostic classification. While benign papillomas can be managed conservatively or with limited excision, papillary carcinomas generally require more extensive surgical intervention, occasionally including sentinel lymph-node biopsy, owing to the potential—though low—risk of nodal metastasis, as described in encapsulated papillary carcinoma (7).

In our cohort, the presence of nuclear atypia and proliferative activity on cytologic examination correlated positively with histologically confirmed neoplasia. Cases with nuclear atypia revealed RON and ROM values of 60% and 20%, respectively, whereas proliferative smears revealed values of 35% and 12%, respectively. These findings parallel previous observations that papillary carcinomas tend to exhibit greater cellularity and more prominent nuclear atypia compared with papillomas (8,9). The

Table 1. Distribution of evaluated cases according to cytologic features and corresponding risk rates

Group	Evaluated cases (n)	Neoplasia (n)	Malignancy (n)	RON (%)	ROM (%)
Nuclear atypia (+)	5	3	1	60.0	20.0
Proliferation (+)	17	6	2	35.3	11.8
Both features (+)	3	2	0	66.7	0
All evaluable cases	25	8	3	32.0	12.0

RON: Risk of neoplasia, ROM: Risk of malignancy

Table 2. Cytologic findings and histologic follow-up of cases with proliferative or atypical features

Findings	Number of cases	Histologic follow-up available	Neoplasia (n)	Malignancy (n)
Nuclear atypia (+)	5	3 (histology) + 2 (clinical follow-up)	3	1 (*in situ* DCIS)
Proliferation (+)	17	6 (histology) + 11 (clinical follow-up)	6	2 (1 IDC, 1 DCIS)
Both features (+)	3	2 (histology) + 1 (clinical follow-up)	2	0
Total	35	8 (histology) + 17 (clinical follow-up)	8	3

IDC: Invasive ductal carcinoma, DCIS: Ductal carcinoma *in situ*

Table 3. Calculated risk metrics for the evaluated cases

Metric	Count	Rate (%)	95% confidence interval
RON	8/25	32.0	17.7–51.6
ROM	3/25	12.0	4.2–30.0

RON: Risk of neoplasia, ROM: Risk of malignancy

combined evaluation of atypia and hypercellularity therefore appears to improve the prediction of neoplastic potential in papillary lesions.

The diagnostic accuracy of FNAC for papillary lesions has been variably reported, ranging from 54% to 88% (4,7). Simsir et al. (6) found that 66% of cytologically "papillary" lesions were benign on excision, but that hypercellularity and nuclear atypia were significant discriminators between papillomas and papillary carcinomas. Similarly, Sauer (7) emphasized that the substantial morphologic overlap between benign and low-grade malignant papillary lesions often necessitates histologic confirmation for definitive classification. Our findings align with these observations and highlight the importance of cautious interpretation of atypical cytologic features.

Immunocytochemistry on cell-block preparations can provide additional diagnostic confidence. Markers such as p63, smooth muscle actin (SMA), and calponin can help identify the presence or absence of a myoepithelial layer and thus aid in distinguishing papillomas from papillary carcinomas. In this setting, p63 demonstrates nuclear staining, whereas SMA and calponin show cytoplasmic staining in myoepithelial cells. Although Reis-Filho et al. (10) demonstrated the diagnostic value of p63 staining in cytologic preparations, no immunohistochemical analyses were applied to the cytologic materials in our study. Nonetheless, in five cases of intraductal papilloma, immunostaining for myoepithelial markers (p63, calponin, and CK5/6) was performed on histologic sections, confirming the presence of a myoepithelial cell layer.

The International Academy of Cytology, Yokohama System for Reporting Breast Cytology, established in 2017, provides a standardized framework for classifying breast FNA samples into five categories: C1 (insufficient), C2 (benign), C3 (atypical), C4 (suspicious), and C5 (malignant) (11). The observed 12% malignancy rate corresponds well with the ROM reported for the indeterminate (C3/C4) categories of the Yokohama System, ranging between 5% and 75% (11). This suggests that the threshold for recommending excision in our series was appropriately conservative. The finding that all histologically sampled cases were neoplastic further reinforces the value of FNAC as a reliable triage tool, especially when interpreted alongside clinical and radiologic findings.

Study Limitations

Several limitations should be acknowledged. The retrospective, single-center design and the limited number of histologically correlated cases reduce the statistical power to assess cytologic-histologic concordance. Moreover, the absence of a standardized scoring system for cytologic atypia introduces potential interobserver variability, a limitation noted in earlier studies (6,9). Despite these constraints, our findings emphasize that combined assessment of nuclear atypia and proliferative activity provides meaningful insight into the neoplastic potential of papillary breast lesions.

Conclusion

FNAC remains a valuable first-line diagnostic modality in the evaluation of papillary breast lesions. Although a definitive cytologic distinction between benign and malignant lesions may not always be achievable, recognizing proliferative and atypical features aids in risk stratification and guides appropriate surgical management. The integration of immunocytochemical studies and standardized reporting frameworks such as the Yokohama System may further enhance diagnostic accuracy and reproducibility in papillary breast cytology.

Ethics

Ethics Committee Approval: Ethical approval for the study was granted by the University of Health Sciences Türkiye, İstanbul Training and Research Hospital Ethics Committee (decision number: 303, date: 05.12.2025).

Informed Consent: Due to the retrospective nature of the study, the requirement for informed consent was waived by the ethics committee.

Footnotes

Authorship Contributions: Surgical and Medical Practices - B.Ö.; Concept - A.B.; Design - B.Ö.; Data Collection or Processing - A.B.; Analysis or Interpretation - B.Ö.; Literature Search - A.B.; Writing - B.Ö.

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. Prathiba D, Rao S, Kshitija K, Joseph LD. Papillary lesions of breast - an introspect of cytomorphological features. J Cytol. 2010; 27: 12-5.
2. Aggarwal D, Soin N, Kalita D, Pant L, Kudesia M, Singh S. Cytodiagnosis of papillary carcinoma of the breast: report of a case with histological correlation. J Cytol. 2014; 31: 119-21.
3. Gomez-Aracil V, Mayayo E, Azua J, Arraiza A. Papillary neoplasms of the breast: clues in fine needle aspiration cytology. Cytopathology. 2002; 13: 22-30.
4. Vijayvergiya G, Naik LP, Kothari KS. Utility of fine needle aspiration cytology (FNAC) in diagnosis of papillary lesions of the breast. Int J Clin Diagn Pathol. 2020; 3: 32-8.
5. Gomes Pinto D, Schmitt FC. Overcoming pitfalls in breast fine-needle aspiration cytology: a practical review. Acta Cytol. 2024; 68: 206-18.
6. Simsir A, Waisman J, Thorner K, Cangiarella J. Mammary lesions diagnosed as "papillary" by aspiration biopsy: 70 cases with follow-up. Cancer. 2003; 99: 156-65.
7. Sauer T. The cytomorphological spectrum of papillary lesions in the breast. M J Cyto. 2017; 1: 005.
8. Michael CW, Buschmann B. Can true papillary neoplasms of breast and their mimickers be accurately classified by cytology? Cancer. 2002; 96: 92-100.
9. Tse GM, Ma TK, Lui PC, Ng DC, Yu AM, Vong JS, et al. Fine needle aspiration cytology of papillary lesions of the breast: how accurate is the diagnosis? J Clin Pathol. 2008; 61: 945-9.
10. Reis-Filho JS, Milanezi F, Amendoeira I, Albergaria A, Schmitt FC. p63 staining of myoepithelial cells in breast fine needle aspirates: a study of its role in differentiating in situ from invasive ductal carcinomas of the breast. J Clin Pathol. 2002; 55: 936-9.
11. Nabat LAAR, Ali HH. A review of the predictive values and malignancy risks of the YOKOHAMA system for reporting breast fine needle aspiration cytology. Int J Clin Diagn Pathol. 2023; 6: 19-23.

İstanbul Med J 2026; 27(2): 159-63
DOI: 10.4274/imj.galenos.2026.49596

Factors Determining Admission to the Intensive Care Unit After Urologic Cancer Surgery and Clinical Outcomes

Özlem Ateşal, Abdurrahman Tünay

University of Health Sciences Türkiye, İstanbul Training and Research Hospital, Clinic of Anesthesiology and Reanimation, İstanbul, Türkiye

ABSTRACT

Introduction: This study aimed to identify the factors associated with postoperative intensive care unit (ICU) admission and to evaluate short-term clinical outcomes in patients undergoing surgery for urological malignancies.

Methods: This retrospective observational study included 75 patients admitted to the ICU following elective surgery for kidney, bladder, or prostate cancer between January 1, 2023, and January 1, 2024. Demographic characteristics, comorbid conditions, and surgery- and anesthesia-related variables were recorded. Reasons for ICU admission, ICU length of stay, and 28-day mortality were analyzed; previously missing 28-day mortality data are now included and reported as 0%. Multivariable regression analysis was omitted because the sample size was small and its application in the previous manuscript was inappropriate.

Results: The mean age of the patients was 66.7±10.9 years, and 84% were male. Kidney cancer (45.3%) was the most common diagnosis, followed by bladder cancer (38.7%) and prostate cancer (16.0%). ICU admission was most frequently performed for postoperative monitoring (53.3%), followed by hemodynamic instability (32.0%) and respiratory failure (14.7%). The median ICU length of stay was 0 days (IQR: 0–1), with 65.3% of patients discharged on the same day.

Conclusion: Most ICU admissions after urological cancer surgery were undertaken for short-term postoperative observation. Associations previously inferred from regression analysis between diabetes mellitus and sex were removed; only univariate differences are reported. These findings highlight the importance of perioperative risk assessment and may contribute to more efficient utilization of intensive care resources.

Keywords: Urological cancer, intensive care unit, postoperative, diabetes mellitus, hemodynamic instability

Introduction

Prostate, bladder, and kidney cancers constitute the most frequently encountered malignancies within urological practice. Prostate cancer is the leading urological malignancy worldwide and remains one of the most commonly diagnosed cancers among men, with marked geographic variation in incidence. Data from GLOBOCAN 2020 indicate that it ranks among the top causes of cancer-related morbidity and mortality in men, and incidence rates are higher in developed countries. In Türkiye, prostate cancer is the second most common malignancy in men and ranks among the leading causes of cancer overall (1). Bladder cancer is also highly prevalent, particularly in male patients, and is listed among the ten most commonly diagnosed cancers globally (2). Renal cell carcinoma represents the most common solid malignancy of the kidney and accounts for approximately 3% of all adult cancers, with a higher incidence in men (3).

The development of urological malignancies is influenced by a combination of genetic predisposition and environmental exposures. Advanced age, tobacco use, obesity, hypertension, chronic kidney disease, occupational exposure to carcinogens, and lifestyle-related factors have all been implicated in the pathogenesis of disease. With increasing age, the prevalence of comorbid conditions such as hypertension, diabetes mellitus (DM), coronary artery disease (CAD), and renal dysfunction also rises. Patients undergoing surgery for urological cancers often have a substantial comorbidity burden.

While some individuals enter the perioperative period with well-controlled chronic illnesses, others may undergo surgery with suboptimal physiological reserve. Management following major oncological surgery requires careful planning, particularly regarding the level of monitoring needed after anesthesia and surgery. Despite widespread use of perioperative risk assessment tools, there is still no universally accepted approach for identifying patients who are most likely to benefit from routine admission to the intensive care unit (ICU) after elective surgery (4). Previous studies have shown that a significant proportion of elderly hospitalized patients require ICU monitoring, and that a considerable share of ICU resources is allocated to postoperative surgical patients (5,6).

Address for Correspondence: Özlem Ateşal, MD, University of Health Sciences Türkiye, İstanbul Training and Research Hospital, Clinic of Anesthesiology and Reanimation, İstanbul, Türkiye
E-mail: ozlemturgut12@gmail.com **ORCID ID:** orcid.org/0000-0003-1473-9415

Cite this article as: Ateşal Ö, Tünay A. Factors determining admission to the intensive care unit after urologic cancer surgery and clinical outcomes. İstanbul Med J. 2026; 27(2): 159-63

Received: 23.03.2026
Accepted: 26.04.2026
Publication Date: 12.05.2026

As surgical techniques and patient complexity increase, the demand for postoperative care environments that allow close observation and early detection of complications has also grown (7,8). The decision to admit a patient to the ICU after surgery is rarely based on a single parameter. Instead, it reflects an integrated evaluation of preoperative patient characteristics, intraoperative events, and the clinical judgment of the anesthesia and surgical teams (9). Appropriate postoperative risk stratification is particularly important in major surgery, as delayed recognition of clinical deterioration has been associated with worse outcomes. Studies have demonstrated increased morbidity and mortality among high-risk surgical patients who required unplanned ICU admission or late transfer from the ward (10-12).

In this context, the present study aimed to investigate the factors determining postoperative ICU admission among patients undergoing elective surgery for kidney, bladder, and prostate cancer, and to evaluate their association with ICU length of stay and 28-day mortality. We hypothesized that postoperative ICU requirement is influenced by preoperative demographic and clinical characteristics, comorbidities, and surgery- and anesthesia-related factors and that these variables may also be associated with short-term clinical outcomes in patients admitted to the ICU.

Primary hypothesis (H1): The need for postoperative intensive care is significantly influenced by patients' preoperative demographic and clinical characteristics, comorbidities, and surgery- and anesthesia-related factors.

Secondary hypothesis (H2): In patients admitted to the ICU, ICU length of stay and 28-day mortality are associated with preoperative characteristics and intraoperative complications.

Methods

This retrospective observational study was conducted in accordance with the principles of the Declaration of Helsinki, after approval had been obtained from the Clinical Research Ethics Committee of University of Health Sciences Türkiye, İstanbul Training and Research Hospital (approval number: 309, date: 19.12.2025). Patients who underwent elective surgery for kidney, bladder, or prostate cancer between January 1, 2023, and January 1, 2024, and who were followed postoperatively in the ICU were included in the study. Patient data were retrospectively reviewed in University of Health Sciences Türkiye, İstanbul Training and Research Hospital information management system and patient follow-up files, then anonymized and recorded.

University of Health Sciences Türkiye, İstanbul Training and Research Hospital has a 37-bed ICU managed by the department of anesthesiology and reanimation and staffed around the clock by specialists and resident physicians. This unit manages critically ill patients and surgical patients requiring close postoperative monitoring.

Inclusion criteria: Patients aged ≥18 years who had undergone elective surgery for kidney, bladder, or prostate cancer, were admitted to the ICU for the first time in relation to the relevant surgical procedure, and were followed in the ICU postoperatively for at least 24 hours; however, the median ICU length of stay may be 0 days in some cases, and this is noted as a limitation.

Exclusion criteria: Patients who underwent emergency surgery, received cardiopulmonary resuscitation in the preoperative or postoperative period, were readmitted to the ICU for the same surgical procedure, had incomplete medical records, or were American Society of Anesthesiologists (ASA) 4-5 (previously not stated consistently, now clarified).

Patients' demographic characteristics (age and sex), comorbid diseases, ASA classifications, type and duration of surgery, anesthesia method used, intraoperative blood transfusion requirement, and complications related to surgery or anesthesia were recorded. Hemodynamic instability was defined as systolic blood pressure <90 mmHg or mean arterial pressure <65 mmHg, requiring vasopressors or fluid resuscitation.

In addition, the reasons for ICU admission were evaluated. ICU admission decisions were based on the clinical assessments of the anesthesiologist and the surgeon; no standardized scoring system was applied, which is a limitation of the study.

Primary endpoint: Determination of the reasons for ICU admission. Secondary endpoints: ICU length of stay and 28-day mortality.

Statistical Analysis

Data were analyzed using IBM SPSS Statistics 26.0 (IBM Corp., Armonk, NY, USA). The conformity of continuous variables to normal distribution was assessed using the Shapiro-Wilk test; variables with normal distribution were expressed as mean ± standard deviation, while those without normal distribution were presented as median and interquartile range (IQR) (median, IQR). Categorical variables were presented as number and percentage [n (%)].

For comparisons between groups, the Kruskal-Wallis test was used for continuous variables, and for variables found to be significant, pairwise comparisons were performed using the Bonferroni-corrected Mann-Whitney U test. Categorical variables were compared using the chi-square test, and Fisher's exact test was used when the expected cell count was less than five.

Multivariable logistic regression analysis, previously included in the manuscript, was removed due to the small sample size and inappropriate application of the method. Only univariate comparisons are presented.

All p values are two-tailed, and $p<0.05$ was considered statistically significant.

Results

Descriptive Characteristics

A total of 75 urological cancer patients admitted to the ICU were included in the study. The mean age of the patients was 66.7±10.9 years (range: 30-90 years), and the majority were male (n=63, 84.0%). Kidney cancer was the most frequent diagnosis (n=34, 45.3%), followed by bladder cancer (n=29, 38.7%), and prostate cancer (n=12, 16.0%).

At least one chronic disease was present in 73.3% of the patients (n=55). Regarding comorbidity distribution, hypertension was observed in 53.3% (n=40), DM in 36.0% (n=27), and CAD in 34.7% (n=26). The majority of patients were classified as ASA II (80.0%); the mean Acute Physiology

and Chronic Health Evaluation II (APACHE II) score was 8.9±4.5. General anesthesia was administered in 81.3% of cases (n=61), and the mean duration of surgery was 219.0±129.0 minutes.

When the reasons for ICU admission were evaluated, 11 patients (14.7%) were admitted for respiratory failure, 24 (32.0%) for perioperative hemodynamic instability, and 40 (53.3%) for postoperative monitoring at the discretion of the anesthesiologist. The median ICU length of stay was 0 days (IQR: 0-1), and 65.3% of patients were discharged on the same day. The 28-day mortality was 0% (Table 1).

Between Groups According to Reason for ICU Admission

No statistically significant differences were found among the three groups in terms of age, APACHE II score, duration of surgery, or ICU length of stay (p=0.307, p=0.282, p=0.336, and p=0.882, respectively). No significant differences were observed among the groups with respect to the presence of at least one chronic disease, cancer type distribution, hypertension, CAD, chronic kidney disease, respiratory disease, endocrine disease, ASA score, or anesthesia type (all p>0.05).

The only statistically significant difference between the groups was observed in the prevalence of DM (p=0.045). The rate of DM was 47.5% in the monitoring group and markedly lower (16.7%) in the hemodynamic instability group. Regarding sex distribution, the difference between groups reached borderline statistical significance (p=0.071); the proportion of women in the monitoring group (25.0%) was higher than that in the other two groups (Table 2).

The results of the regression analysis are summarized in Table 3.

Discussion

This study examined postoperative ICU admissions following surgery for urological malignancies, focusing on factors associated with indications for ICU admission and with short-term outcomes. ICU admissions were most commonly for postoperative surveillance rather than for management of acute complications. ICU length of stay was short, and the majority of patients were discharged on the same day.

The demographic characteristics of the study population are consistent with the known epidemiological distribution of urological cancers (1-3). The predominance of male patients and the higher mean age reflect the age- and sex-related distribution of these malignancies. In addition, the high prevalence of comorbid diseases, particularly hypertension, DM, and CAD, indicates that patients undergoing urological oncological surgery generally represent an older population with a high comorbidity burden.

DM was associated with ICU admission for monitoring rather than for hemodynamic instability, likely reflecting cautious perioperative management rather than a protective physiological effect. Similarly, differences observed between sexes are limited by the small number of females. These associations should be interpreted cautiously.

No significant differences were found between groups with respect to age, APACHE II score, duration of surgery, ASA score, or ICU length of stay. This suggests that classical risk indicators alone cannot explain ICU admission decisions; clinical judgment and institutional protocols are likely to influence them.

Comparison with previous studies shows similar patterns of ICU utilization (13-15). ICU admission is often based on anticipated monitoring needs rather than active management of complications.

Study Limitations

First, the study had a retrospective, single-center design. Data were obtained from hospital records, and some clinical parameters may have been missing or recorded inconsistently.

Second, the relatively small sample size (n=75) reduced statistical power, particularly for subgroup comparisons. Because the number of patients admitted to the ICU due to respiratory failure was low, no multivariable model could be established for this group.

Table 1. Demographic and clinical characteristics of the patients

Variable	All patients (n=75)
Age (years), mean ± SD (min-max)	66.7±10.9 (30–90)
Sex	
Male	63 (84.0%)
Female	12 (16.0%)
Cancer type	
Kidney cancer	34 (45.3%)
Bladder cancer	29 (38.7%)
Prostate cancer	12 (16.0%)
Comorbid diseases	
At least one chronic disease	55 (73.3%)
Hypertension	40 (53.3%)
Diabetes mellitus	27 (36.0%)
Coronary artery disease	26 (34.7%)
Chronic kidney failure	2 (2.7%)
Respiratory disease	13 (17.3%)
Endocrine disease	4 (5.3%)
ASA score	
ASA I	4 (5.3%)
ASA II	60 (80.0%)
ASA III	11 (14.7%)
APACHE II score, mean ± SD (min–max)	8.9±4.5 (0–32)
Anesthesia type	
General anesthesia	61 (81.3%)
Spinal anesthesia	14 (18.7%)
Duration of surgery (min), mean ± SD (min-max)	219.0±129.0 (20–595)
ICU length of stay (days), median (IQR)	0.0 (0.0–1.0)
Reason for ICU admission	
Respiratory failure	11 (14.7%)
Hemodynamic instability	24 (32.0%)
Monitoring purpose	40 (53.3%)

Mean: Average, SD: Standard deviation; IQR: Interquartile range; ASA: American Society of Anesthesiologists; APACHE II: Acute Physiology and Chronic Health Evaluation II; ICU: Intensive care unit, min-max: Minimum, maximum

Table 2. Comparison between groups according to reason for ICU admission				
Variable	**Group 1 respiratory failure (n=11)**	**Group 2 hemodynamic instability (n=24)**	**Group 3 monitoring purpose (n=40)**	**p**
Continuous variables-mean ± SD/median (IQR)				
Age (years)	70.8±8.3	66.5±12.9	65.7±10.1	0.307
APACHE II score	11.8±7.6	8.7±3.6	8.2±3.6	0.282
Duration of surgery (min)	178.2±118.9	205.6±143.5	238.2±121.8	0.336
ICU length of stay (days), median (IQR)	0.0 (0.0–1.0)	0.0 (0.0–1.0)	0.0 (0.0–1.0)	0.882
Sex, n (%)				
Male	10 (90.9%)	23 (95.8%)	30 (75.0%)	0.071
Cancer type, n (%)				
Kidney cancer	6 (54.5%)	9 (37.5%)	19 (47.5%)	0.547
Bladder cancer	5 (45.5%)	10 (41.7%)	14 (35.0%)	
Prostate cancer	0 (0.0%)	5 (20.8%)	7 (17.5%)	
Comorbid diseases, n (%)				
At least one chronic disease	9 (81.8%)	16 (66.7%)	30 (75.0%)	0.604
Hypertension	6 (54.5%)	11 (45.8%)	23 (57.5%)	0.661
Diabetes mellitus	4 (36.4%)	4 (16.7%)	19 (47.5%)	**0.045**
Coronary artery disease	3 (27.3%)	9 (37.5%)	14 (35.0%)	0.838
Chronic kidney failure	0 (0.0%)	0 (0.0%)	2 (5.0%)	0.407
Respiratory disease	2 (18.2%)	5 (20.8%)	6 (15.0%)	0.834
Endocrine disease	0 (0.0%)	1 (4.2%)	3 (7.5%)	0.590
ASA score, n (%)				
ASA I	0 (0.0%)	3 (12.5%)	1 (2.5%)	0.229
ASA II	9 (81.8%)	16 (66.7%)	35 (87.5%)	
ASA III	2 (18.2%)	5 (20.8%)	4 (10.0%)	
Perioperative data, n (%)				
General anesthesia	8 (72.7%)	20 (83.3%)	33 (82.5%)	0.728

Statistical tests used: Kruskal-Wallis test (continuous variables); chi-square test (categorical variables). A p value of <0.05 was considered statistically significant. Significant p value are shown in **red**. SD: Standard deviation; IQR: Interquartile range; ASA: American Society of Anesthesiologists; APACHE II: Acute Physiology and Chronic Health Evaluation II; ICU: Intensive care unit, min: Minute

Table 3. Multinomial logistic regression analysis-hemodynamic instability vs. monitoring purpose			
Variable	**aOR**	**95% confidence interval**	**p**
Hemodynamic instability vs monitoring purpose (reference)			
Diabetes mellitus	0.37	0.15–0.85	0.016
Sex (male)	0.39	0.20–0.88	
Model fit statistics			
Nagelkerke R^2	0.178		
AUC (ROC)	0.689		
Likelihood ratio χ^2	8.975 (df=2)		

aOR: Adjusted odds ratio, AUC: Area under the curve, ROC: Receiver operating characteristic

Third, only patients admitted to the ICU were included. Patients undergoing the same procedures but who were not admitted were excluded, thereby limiting the ability to compare determinants of ICU need in the broader surgical population.

Fourth, ICU admission decisions were based on the judgment of anesthesiologists and surgeons rather than on standardized criteria, which may have influenced the high proportion of admissions for monitoring.

Fifth, the inclusion criterion of "at least 24 hours in ICU" is inconsistent with the median ICU stay of 0 days and with 65% of patients being discharged the same day; this discrepancy represents a limitation.

Finally, long-term outcomes (e.g., 90-day mortality and readmission) were not evaluated.

Conclusion

This retrospective study demonstrates that the majority of ICU admissions following urological cancer surgery were for short-term postoperative monitoring. Associations previously inferred from regression analysis regarding DM and sex were removed; only univariate differences are reported. The clinical significance of these associations should be confirmed in larger, multicenter studies.

In a population undergoing elective surgery, excluding ASA 4-5 patients, short ICU stays and low mortality underscore the importance of perioperative risk assessment to ensure efficient use of intensive care

resources. Prospective, multicenter studies may contribute to a more comprehensive evaluation of perioperative risk factors and optimization of ICU resource utilization.

Ethics

Ethics Committee Approval: This retrospective observational study was approved by the Clinical Research Ethics Committee of University of Health Sciences Türkiye, İstanbul Training and Research Hospital (approval number: 309, date: 19.12.2025).

Informed Consent: Patient data were retrospectively reviewed in University of Health Sciences Türkiye, İstanbul Training and Research Hospital information management system and patient follow-up files, then anonymized and recorded.

Footnotes

Authorship Contributions: Surgical and Medical Practices - Ö.A., A.T.; Concept - Ö.A., A.T.; Design - Ö.A., A.T.; Data Collection or Processing - Ö.A., A.T.; Analysis or Interpretation - Ö.A., A.T.; Literature Search - Ö.A., A.T.; Writing - Ö.A., A.T.

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. Abouassaly R, Thompson IM Jr, Platz EA, Klein EA. Prostate cancer. In: Campbell-Walsh Urology. 10th ed. p. 2704-26.
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021; 71: 209-49.
3. Capitanio U, Bensalah K, Bex A, Boorjian SA, Bray F, Coleman J, et al. Epidemiology of renal cell carcinoma. Eur Urol. 2019; 75: 74-84.
4. Pearse RM, Harrison DA, James P, Watson D, Hinds C, Rhodes A, et al. Identification and characterisation of the high-risk surgical population in the United Kingdom. Crit Care. 2006; 10: R81.
5. Yu W, Ash AS, Levinsky NG, Moskowitz MA. Intensive care unit use and mortality in the elderly. J Gen Intern Med. 2000; 15: 97-102.
6. Sülekli HE, Küçük A. Intensive care units research. Republic of Türkiye Ministry of Health, Turkish Public Hospitals Institution; 2015. p. 1-7.
7. Squadrone V, Coha M, Cerutti E, Schellino MM, Biolino P, Occella P, et al. Continuous positive airway pressure for treatment of postoperative hypoxemia: a randomized controlled trial. JAMA. 2005; 293: 589-95.
8. Pearse R, Dawson D, Fawcett J, Rhodes A, Grounds RM, Bennett ED. Early goal-directed therapy after major surgery reduces complications and duration of hospital stay. A randomised, controlled trial [ISRCTN38797445]. Crit Care. 2005; 9: R687-93.
9. Vimlati L, Gilsanz F, Goldik Z. Quality and safety guidelines of postanaesthesia care: Working Party on Post Anaesthesia Care (approved by the European Board and Section of Anaesthesiology, Union Européenne des Médecins Spécialistes). Eur J Anaesthesiol. 2009; 26: 715-21.
10. NELA Project Team. Third Patient Report of the National Emergency Laparotomy Audit. London: RCoA; 2017.
11. The Royal College of Surgeons of England, Department of Health. The Higher Risk General Surgical Patient: Towards Improved Care for a Forgotten Group. London: Department of Health; 2011.
12. Gillies MA, Harrison EM, Pearse RM, Garrioch S, Haddow C, Smyth L, et al. Intensive care utilization and outcomes after high-risk surgery in Scotland: a population-based cohort study. Br J Anaesth. 2017; 118: 123-31.
13. Ohbe H, Matsui H, Kumazawa R, Yasunaga H. Postoperative ICU admission following major elective surgery: a nationwide inpatient database study. Eur J Anaesthesiol. 2022; 39: 436-44.
14. Efe S, Sak İ, İnal V. A retrospective analysis of quality indicators in a mixed-type tertiary center intensive care unit. J Crit Intensive Care. 2018; 9: 1-6.
15. Coşkun AS. One-year evaluation of patients admitted to postoperative intensive care. Acta Med Nicomedia. 2020; 3: 105-8.

İstanbul Med J 2026; 27(2): 164-70
DOI: 10.4274/imj.galenos.2026.65657

The Impact of a Structured Homecare Intervention Program on Caregivers' Knowledge of Children with Cerebral Palsy in Kirkuk City

Huda Faeeq Mohammed, Shukir Saleem Hasan

Pediatric Nursing Department, College of Nursing, Hawler Medical University, Erbil, Iraq

ABSTRACT

Introduction: Cerebral palsy (CP) describes a group of long-term neurological disorders that affect posture, mobility, and muscle tone due to damage to the developing brain, usually before or immediately after birth. Children with CP may have difficulty with balance, coordination, and fine motor skills. Some may also experience problems with speech, vision, or learning. The severity of symptoms can vary widely between people. For children with CP, prompt diagnosis and treatment—such as occupational therapy, physical therapy, and supportive care—are essential to improving function, independence, and general quality of life.

Methods: A quasi-experimental study design was used. Caregivers and their children who had CP attended physiotherapy in the rehabilitation units at Azadi Teaching Hospital in Kirkuk City, Iraq. Non-probability sampling (purposive) was used to recruit 60 caregivers (30 in the intervention group and 30 in the control group). The intervention group received a specific educational program to improve caregivers' knowledge of home care, while the control group did not receive it. Only caregivers caring for children with CP were included. Data were collected through face-to-face interviews; the Mann-Whitney test was used to compare pre- and post-test measurements. The p value ≤0.05 was considered a statistically significant difference.

Results: Most caregivers (66.7-76.7%) were mothers from diverse educational and socioeconomic backgrounds. The most frequent type of spastic CP, occurring in 66.7-76.7% of cases, was observed in both groups. The pre-test results revealed no significant difference between the two groups before implementing the program. The program's effectiveness was demonstrated by the intervention group's statistically significant difference between pre- and post-test scores, whereas the control group's post-test scores remained unchanged.

Conclusion: The structured educational program significantly enhanced caregivers' knowledge in the intervention group, with no improvement in the control group.

Keywords: Caregivers, cerebral palsy, program evaluation, delivery of healthcare

Introduction

Cerebral palsy (CP) is a long-term, non-progressive disorder affecting the developing brain of the fetus or newborn. It's one of the most common causes of physical disability and is associated with motor disorders, causing problems with perception, sensation, cognition, and communication (1). Historically, CP was first described as a movement disorder in ancient Sumerian and Hippocratic writings (2). The disorder results from brain damage during a critical stage of cerebral cortex development (3). Such an injury may occur either before birth or during the child's first five years of life. Several factors such as premature birth, stroke, birth asphyxia, genetic abnormalities, head injuries, home delivery, hypoxic-ischemic encephalopathy, intrauterine growth restriction, hyperbilirubinemia, low birth weight, microcephaly, multiple pregnancies, small for gestational age, infections (such as chorioamnionitis, maternal urinary tract infections, cytomegalovirus, and neurotropic virus infection), and other factors can lead to CP (4,5). Globally, this neuro-developmental disorder affects around 17 million people (5). The prevalence of CP ranges from 1 to 4 per 1000 live births worldwide, predominantly reported in high-income countries, while data from low- and middle-income countries suggest higher rates, with estimates of 3 to 4 per 1,000 live births (6). The incidence of CP is estimated at approximately 3 per 1,000 children, with a reported male-to-female ratio of 1.4:1 (7). Classification of CP is primarily based on the predominant type of motor impairment, which may include involuntary

Address for Correspondence: Huda Faeeq Mohammed, MD, Pediatric Nursing Department, College of Nursing, Hawler Medical University, Erbil, Iraq
E-mail: hudafaeeq@gmail.com **ORCID ID:** orcid.org/0009-0001-2484-2128

Cite this article as: Faeeq Mohammed H, Saleem Hasan S. The impact of a structured homecare intervention program on caregivers' knowledge of children with cerebral palsy in Kirkuk City. İstanbul Med J. 2026; 27(2): 164-70

Received: 06.11.2025
Accepted: 27.04.2026
Publication Date: 12.05.2026

movements such as dyskinesia, spasticity, and ataxia (8). According to topographical distribution, CP is further categorized into diplegia, quadriplegia, hemiplegia, triplegia, and monoplegia (9-12).

The most frequent complications associated with CP are related to movement, self-care, and communication. Social, intellectual, and physical isolation are more prevalent among children diagnosed with CP (13). Based on the severity of impairments, children with CP can experience issues such as spasticity, contractures, scoliosis, joint dislocations, skin breakdown, aspiration pneumonia, malnutrition, dental issues, low bone mass, epilepsy, psychological challenges, poor vision, perceptual impairment, difficulties while performing self-care tasks like eating, dressing, getting around, and bathing, and complications with bowel and bladder (1,14). The Management of CP requires professional medical, therapeutic, social, and educational programs because there is no widely recognized cure. Physical therapy, occupational therapy, speech therapy, behavioral therapy, medication, and surgery play essential roles in patient care (1,15). Botulinum toxin type A (Botox) may lessen spasticity, and the treatment of dystonia has historically involved anti-Parkinsonian drugs, such as anticholinergic and dopaminergic drugs, and anti-spasticity agents such as baclofen. However, anticonvulsants, anti-dopaminergic drugs, and antidepressants have also been tested (16). Beyond medical treatment, caring for children with CP requires a comprehensive home-care approach in which caregivers play a crucial role. Therefore, assessing caregivers' knowledge and practices is essential to ensure effective care and optimal outcomes for affected children (1,17).

Ensuring enough nourishment, protecting skin health and safety, providing emotional support, promoting growth and development, and instructing parents on how to care for their child are all important aspects of caring for children with CP (18,19). Parents who care for their children with CP play a critical role in helping those children achieve significant developmental milestones. Caregivers who have a better understanding of CP are better equipped to handle caregiving duties; well-informed caregivers are more likely to have high self-efficacy and positive psychosocial indicators and to offer enhanced care and support. The health and quality of life of children with CP can be enhanced by caregivers by improving their understanding of the illness, which often leaves them ill-prepared to confidently provide the care their children require (1,14,20). Insufficient parental awareness of CP can exacerbate disorders, hindering children's growth, while well-informed individuals promote a cooperative mindset, indirectly benefiting their children's development (21). However, caregivers of children with CP frequently face considerable physical and psychological challenges, including stress and health-related problems, which may restrict their capacity to offer appropriate care (22). Therefore, nurse-prepared, structured educational programs are crucial for enhancing caregivers' knowledge and ability to manage children with CP at home. These initiatives assist parents by providing guidance on appropriate nutrition, physical well-being, care techniques, and engagement activities, while also supporting them in building confidence and skills in everyday caregiving. Previous studies have demonstrated that caregivers involved in nurse-led structured initiatives show a markedly improved understanding of CP management and deliver enhanced care quality to their children (15,23).

Objectives of the study

- Find out the socio-demographical characteristics of caregivers.
- Find out the biographical information of children with CP in Kirkuk City.
- Assess the level of caregiver's knowledge regarding home-care pre and post-tests.
- Find out the impact of home-care program on caregivers' knowledge pre and post-tests.
- Find out the association between selected socio-demographic characteristics of the study sample and caregivers knowledge.
- Construct an educational program for caregivers' who take care of children with CP.

Methods

A quasi-experimental study was conducted among 60 caregivers and their children with CP who received physiotherapy in the rehabilitation unit of Azadi Teaching Hospital, Kirkuk City, Iraq. The study period was from May 1st to August 1st, 2025. Participants were selected using a non-probability (purposive) sampling technique because of the exacting inclusion criteria and the small number of caregivers who visited the rehabilitation center throughout the research period. They were divided into two groups: the intervention group consisted of thirty caregivers who were exposed to the educational program, and the control group consisted of thirty caregivers who were not exposed to the program. Additionally, the educational program was developed based on relevant literature and focused on caregivers' knowledge of CP and home-care management, including general information, causes, signs and symptoms, complications, treatment, and critical home-care practices. The program was provided through seven structured face-to-face sessions conducted by the researcher, each lasting approximately 30 minutes, over two weeks. For data collection, an appropriate tool was developed based on an extensive review of the literature and divided into two parts. Part I included two subsections. Section I focused on the socio-demographic characteristics of caregivers, such as age, gender, and socioeconomic status. Section II presents the demographic characteristics of children with CP, providing details such as age and sex. Part II: Assessment of caregiver knowledge about CP. It comprises 38 items designed to assess knowledge. Dichotomous scoring was used: one (1) for a correct answer and zero (0) for an incorrect answer. The pre-test was performed immediately before the intervention, and the post-test was given two weeks after the intervention group had completed the instructional program. Verbal consent was received from all caregivers. Voluntary participation was assured; participants were free to withdraw at any time without penalty; privacy was maintained; confidentiality was ensured; and the study was approved by the Ethical Committee of the Hawler Medical University, College of Nursing (approval number:

2425, date: 22.08.2024) before starting the study. Content validity was established by a panel of experts, and Cronbach's alpha (reliability coefficient) was 0.77.

Statistical Analysis

The data were analyzed using the Statistical Package for Social Sciences (SPSS version 27). In order to examine the differences between the two independent groups, the Mann-Whitney U test used as the Shapiro-Wilk test revealed that the data were not normally distributed ($p<0.05$). A p value of less than 0.05 was deemed statistically significant.

Demographic Characteristics of Caregivers

Most caregivers of children with CP were mothers (76.7% and 66.7% in the control and intervention groups, respectively). Most of the studied mothers were aged 35 years and above, represented by 53.3% of the control group and 46.7% of the intervention group. Regarding education, 36.7% of caregivers in the control group were illiterate, while 30% of caregivers in the intervention group could read and write. Most caregivers were housewives in the control and intervention groups (70% and 73.3%, respectively). The majority had full-term children (70% vs. 60%), and most had only one affected child in the family (86.7% vs. 100%) (Table 1).

Table 1. Distribution of socio-demographic characteristics of caregivers of CP children

	Variables	Control group F. (%)	Intervention group F. (%)
The caregiver	Mother	**23 (76.7)**	**20 (66.7)**
	Father	4 (13.3)	4 (13.3)
	Other	3 (10)	6 (20)
Age; years old	21-27	1 (3.3)	10 (33.3)
	28-34	13 (43.3)	6 (20)
	35-41 and more	**16 (53.3)**	**14 (46.7)**
Level of education	Illiterate	**11 (36.7)**	5 (16.7)
	Can read and write	6 (20)	**9 (30)**
	Primary school graduate	6 (20)	6 (20)
	Secondary school graduate	1 (3.3)	6 (20)
	Diploma or college graduate	6 (20)	3 (10)
	Postgraduate certificate	0 (0)	1 (3.3)
Occupation	Public employee	6 (20)	3 (10)
	Private employee	1 (3.3)	0 (0)
	Self-employee	2 (6.7)	4 (13.3)
	Student	0 (0)	1 (3.3)
	Housewife	**21 (70)**	**22 (73.3)**
Gestational age	Full-term	**21 (70)**	**18 (60)**
	Pre-term	9 (30)	11 (36.7)
	Post-term	0 (0)	1 (3.3)
Number of affected CP children	One child	**26 (86.7)**	**30 (100)**
	Two children	3 (10)	0 (0)
	3 and more children	1 (3.3)	0 (0)

CP: Cerebral palsy

Distribution of Biographical Characteristics of CP Children

In the control group, 50% of children were at the preschool developmental stage, compared with 56.7% of children in the intervention group who were mainly at the toddler stage. Among the CP children, 56.7% of the control group were female, while the intervention group had equal proportions of males and females. Regarding birth weight, 60% of CP children were underweight, whereas 43.3% were of normal weight. Most children in both groups were born in public hospitals (90% vs. 73.3%). The 53.3% of children in the control group cried immediately after birth, whereas 73.3% of children in the intervention group did not cry. The most frequent type of spastic CP was observed in both groups (66.7-76.7%) (Table 2).

Compare the Pre-Test of Both Groups (Control and Intervention) Regarding Mothers' Knowledge

Mothers' pre-intervention knowledge in both the control and intervention groups was assessed to determine baseline similarities before implementing the structured home care program; this assessment revealed no significant differences between the groups. General knowledge had the highest p value ($p=0.692$), followed by knowledge of causes ($p=0.064$); the mean difference was -0.80. Signs

Table 2. Distribution of biographical characteristics of CP children (control and intervention groups) n=60

	Variables	Control group F. (%)	Intervention group F. (%)
Developmental stage	Toddler	11 (36.7)	**17 (56.7)**
	Pre-school age	**15 (50)**	10 (33.3)
	School age	4 (16.7)	3 (10)
Gender	Male	13 (43.3)	**15 (50)**
	Female	**17 (56.7)**	**15 (50)**
Weight at time of birth	Less than 1,000 g	4 (13.3)	4 (13.3)
	Less than 1,500 g	0 (0)	5 (16.7)
	Less than 2,500 g	**18 (60)**	7 (23.3)
	2500 g-3999 g	8 (26.7)	**13 (43.3)**
	High birth weight (>4000 g)	0 (0)	1 (3.3)
Place of birth	In general hospital	**27 (90)**	**22 (73.3)**
	In private hospital	1 (3.3)	5 (16.7)
	At home	2 (6.7)	3 (10)
Cry of the child at birth	Soon after birth	**16 (53.3)**	8 (26.7)
	After some time	14 (46.7)	**22 (73.3)**
Type of CP	Spastic	**23 (76.7)**	**20 (66.7)**
	Athetoid	1 (3.3)	1 (3.3)
	Ataxic	0 (0)	2 (6.7)
	Mixed	6 (20)	7 (23.3)

CP: Cerebral palsy

and symptoms had a mean difference of -0.67 (p=0.083); problems had a mean difference of -0.47 (p=0.278) with the largest upper confidence interval (CI) (0.39); and therapy had a mean difference of -0.27 (p=0.290). Overall, mothers' knowledge showed the largest mean difference (-2.07; p=0.074; 95 % CI: -4.34 to 0.21) indicating that the two groups were similar in knowledge before the implementation of the program (Table 3).

Comparison Between Pre-Test and Post-Test of the Control Group Regarding Mothers' Knowledge

The effect of the health education program on mothers' knowledge in the control group was assessed by comparing pre-test and post-test results. The analysis revealed no significant differences in knowledge domains between the pre- and post-tests. The p values for general knowledge, knowledge of causes, and signs and symptoms were p=0.716 (upper CI: 0.862), p=0.539 (upper CI: 0.522), and p=0.938 (upper CI: 0.819), respectively. Knowledge of difficulties had p=0.714 (upper CI: 0.858), whereas therapy showed the highest t value (1.157) with a mean difference of 0.333 (upper CI: 0.910). Overall, mothers' knowledge showed p=0.777, a mean difference of 0.333, and the broadest CI (-2.01 to 2.68), indicating that knowledge levels remained similar in the control group following the program (Table 4).

Comparison Between Pre-Test and Post-Test of the Intervention Group Regarding Mothers' Knowledge

The impact of the health education program on mothers' knowledge in the intervention group was evaluated by comparing pre-test and post-test results. The analysis showed statistically significant improvements

Table 3. Comparison between pre-test of both groups (control and intervention) group regarding mothers' knowledge

Knowledge	Domain	T	Df	Sig. (2-tailed)	Mean difference	Std. error difference	95% confidence interval of the difference	
							Lower	Upper
General knowledge	Control pre-test	0.399	58	0.692	0.13333	0.33448	-0.53620	0.80287
	Intervention pre-test	0.399	51.560	0.692	0.13333	0.33448	-0.53799	0.80465
Knowledge on causes	Control pre-test	-1.887	58	0.064	-0.80000	0.42390	-1.64853	0.04853
	Intervention pre-test	-1.887	57.744	0.064	-0.80000	0.42390	-1.64861	0.04861
Knowledge on sign and symptoms	Control pre-test	-1.764	58	0.083	-0.66667	0.37794	-1.42319	0.08985
	Intervention pre-test	-1.764	56.720	0.083	-0.66667	0.37794	-1.42355	0.09022
Knowledge on difficulties*	Control pre-test	-1.096	58	0.278	-0.46667	0.42571	-1.31881	0.38548
	Intervention pre-test	-1.096	54.449	0.278	-0.46667	0.42571	-1.32000	0.38666
Knowledge on treatment	Control pre-test	-1.068	58	0.290	-0.26667	0.24975	-0.76660	0.23326
	Intervention pre-test	-1.068	56.557	0.290	-0.26667	0.24975	-0.76687	0.23354
Overall mothers' knowledge	Control pre-test	-1.818	58	0.074	-2.06667	1.13678	-4.34217	0.20884
	Intervention pre-test	-1.818	57.196	0.074	-2.06667	1.13678	-4.34285	0.20952

*Difficulties include: drooling, difficulty in swallowing, functional constipation, epilepsy, sensory defect, stiff joints, dental caries, and delayed speech millstone). Sig.: Significant, Std.: Standard, Df: Degrees of freedom

Table 4. Comparison between control (pre-test and post-test) group before, and after implement health education program regarding mother's knowledge

Compare between mothers' knowledge after implementation of education program		T	Df	Sig. (2-tailed)	Mean difference	Std. error difference	95% confidence interval of the difference	
							Lower	Upper
General knowledge	Control pre-test	0.366	58	0.716	0.13333	0.36410	-0.59549	0.86215
	Control post-test 1	0.366	56.850	0.716	0.13333	0.36410	-0.59580	0.86247
Knowledge on causes	Control pre-test	-0.619	58	0.539	-0.23333	0.37717	-0.98833	0.52166
	Control post-test 1	-0.619	56.197	0.539	-0.23333	0.37717	-0.98885	0.52218
Knowledge on sign and symptoms	Control pre-test	-0.078	58	0.938	-0.03333	0.42566	-0.88539	0.81872
	Control post-test 1	-0.078	57.500	0.938	-0.03333	0.42566	-0.88555	0.81888
Knowledge on difficulties*	Control pre-test	0.368	58	0.714	0.13333	0.36209	-0.59147	0.85814
	Control post-test 1	0.368	57.951	0.714	0.13333	0.36209	-0.59149	0.85815
Knowledge on treatment	Control pre-test	1.157	58	0.252	0.33333	0.28821	-0.24358	0.91025
	Control post-test 1	1.157	57.048	0.252	0.33333	0.28821	-0.24379	0.91045
Overall mothers' knowledge	Control pre-test	0.285	58	0.777	0.33333	1.17163	-2.01194	2.67861
	Control post-test 1	0.285	57.837	0.777	0.33333	1.17163	-2.01208	2.67875

*Difficulties include: drooling, difficulty in swallowing, functional constipation, epilepsy, sensory defect, stiff joints, dental caries, and delayed speech millstone) Sig.: Significant, Std.: Standard, Df: Degrees of freedom

in all knowledge domains following the program (p=0.000). The greatest increase was observed in total knowledge (mean difference: 15.902, 95% CI: 14.32-17.49), followed by an increase in knowledge of causes (mean difference: 7.767, t: 25.09). Other areas showed substantial improvement, including general knowledge (mean difference: 1.581), signs and symptoms (mean difference: 3.800), problems (mean difference: 3.367), and therapy (mean difference: 2.550), indicating the effectiveness of the program in enhancing mothers' understanding (Table 5).

Discussion

The present study examined the demographic characteristics of caregivers and children with CP, as well as caregivers' knowledge before and after an educational intervention, and compared the findings with previous literature.

Mothers constituted the largest proportion of caregivers in both groups, accounting for 76% of caregivers in the control group and 66% of caregivers in the intervention group. Fathers accounted for 3-13% in each group, whereas other family members provided more care in the intervention group (20%) than in the control group (10%). Almosallam et al. (1) and Hussein et al. (19) found comparable results, showing that mothers are the major caregivers for daily physical care, medical appointments, and treatment adherence. Bibi et al. (22) and Abo Hamed (23) confirmed that mothers play the primary caring role for children with CP in a variety of cultural situations. 46.7% of caregivers in the intervention group and 53.3% in the control group were 35 years of age or older. Thirty-three percent of caregivers in the intervention group were younger than caregivers in the control group. The age range of these caregivers was 21 to 27 years, which differs from the average ages reported in other studies (36.35±6.97 years and 36.76±7.85 years). As a result of the long-term caregiving responsibilities for children with CP, previous studies by Almosallam et al. (1), Rashad et al. (12), Gad Ahmed et al. (15), Bibi et al. (22), Hamed and Abo Hamed (23), Ramadan Mohamed Ebeed et al. (24), Samia et al. (25), Gamal et al. (26), and Jahan et al. (27), have found that most caregivers are in their early to middle years of life.

Regarding employment, the percentage of caregivers who were housewives in the present study was between 21% and 22%, substantially lower than reported in other studies (for instance, 75% to 80%). Cultural norms, socioeconomic conditions, and sampling methods may explain this difference [Almosallam et al. (1); Rashad et al. (12); Hussein et al. (19); Samia et al. (25); Gamal et al. (26); Jahan et al. (27)]. Regarding education, the present study showed an 11% illiteracy rate in the control group and a 9% literacy rate for those who could "read and write" in the intervention group, consistent with the wide variety in educational levels discovered in prior studies, from high educational level of 53.7% to significant rates of illiteracy. These demographic discrepancies are significant because caregivers' ability to provide care and their access to essential resources are affected by their educational background and employment status. According to earlier studies by Almosallam et al. (1), Rashad et al. (12), Gad Ahmed et al. (15), Bibi et al. (22), Samia et al. (25), Gamal et al. (26), Jahan et al. (27), caregivers' employment status and educational background have a direct impact on their capacity to manage CP, provide care, and make effective use of healthcare resources.

In the current sample, 50% of the children in the control group were preschool-aged, similar to previous research, whereas 56.7% of the intervention group were toddlers, and there were fewer school-aged children in both groups. On the other hand, the median age of the children was 8 years. The control group had a higher proportion of females (56.7%) than males, but the intervention group had equal gender distribution. Gender distribution has been reported to vary across studies. Low birth weight (<2500 g) was more common in the control group (73.3%) than in the intervention group (43.3%), whereas a

Table 5. Comparison between intervention (pre-test and post-test) groups after implement health education program regarding mother's knowledge

Compare between pre and post-test of intervention after implement the education program		T	Df	Sig. (2-tailed)	Mean difference	Std. error difference	95% confidence interval of the difference	
							Lower	Upper
General knowledge	Intervention pre-test	-7.384	56	0.000	-1.58095	0.21411	-2.00987	-1.15203
	Intervention post-test 1	-7.560	40.488	0.000	-1.58095	0.20911	-2.00342	-1.15848
Knowledge on causes	Intervention pre-test	24.225	56	0.000	7.76667	0.32061	7.12441	8.40892
	Intervention post-test 1	25.090	29.000	0.000	7.76667	0.30955	7.13356	8.39978
Knowledge on sign and symptoms	Intervention pre-test	14.893	56	0.000	3.80000	0.25515	3.28887	4.31113
	Intervention post-test 1	15.425	29.000	0.000	3.80000	0.24635	3.29615	4.30385
Knowledge on difficulties*	Intervention pre-test	9.638	56	0.000	3.36667	0.34932	2.66689	4.06644
	Intervention post-test 1	9.982	29.000	0.000	3.36667	0.33728	2.67686	4.05647
Knowledge on treatment	Intervention pre-test	13.712	56	0.000	2.55000	0.18597	2.17745	2.92255
	Intervention post-test 1	14.005	43.151	0.000	2.55000	0.18208	2.18285	2.91715
Overall mothers' knowledge	Intervention pre-test	20.124	56	0.000	15.90238	0.79021	14.31941	17.48535
	Intervention post-test 1	20.809	30.475	0.000	15.90238	0.76420	14.34269	17.46207

*Difficulties include: drooling, difficulty in swallowing, functional constipation, epilepsy, sensory defect, stiff joints, dental caries, and delayed speech millstone).
Sig.: Significant, Std.: Standard, Df: Degrees of freedom

higher proportion of high birth weight (>4000 g) infants was observed in the intervention group. This indicates differing perinatal characteristics, resembling those observed in previous studies. Rashad et al. (12), Gad Ahmed et al. (15), Hussein et al. (19), Samia et al. (25), and Jahan et al. (27) have previously documented that the majority of children with CP are diagnosed in early childhood, that the gender distribution varies among studies, and that low and high birth weights are significant risk factors during the perinatal period. Most children in both groups were born in general hospitals (90% vs. 73.3%), although the intervention group had a slightly higher percentage of deliveries in private hospitals and home births. More than half of the children in the control group cried immediately (53.3%), while fewer children in the intervention group did so (26.7%). In the present study, birth histories showed between-group variation in the timing of the first cry, similar to observations in other research. In all groups, spastic CP was the most prevalent subtype (76.7% and 66.7%), followed by the mixed type, with one occurrence of ataxic CP (6.7%), which is consistent with epidemiological statistics from comparable studies.

The current study reveals no statistically significant differences in any of the knowledge domains between the mothers in the study's control and intervention groups' baseline knowledge before implementing the educational program ($p>0.05$), with the greatest p value (0.692) found in general knowledge, followed by knowledge of causes ($p=0.064$), knowledge of signs and symptoms ($p=0.083$), knowledge of difficulties ($p=0.278$), knowledge of treatment ($p=0.290$), and overall mothers' knowledge ($p=0.074$). These results indicate homogeneity between groups at baseline, consistent with the baseline similarities reported in previous studies. Analysis of pre- and post-program knowledge levels in the control group showed no statistically significant differences ($p>0.05$) across general knowledge, causes, signs and symptoms, difficulties, and treatment; the mean difference was only 0.33 ($p=0.777$). This suggests that mothers' understanding did not improve in the absence of the educational intervention. These results align with previous studies indicating that caregivers' knowledge remains largely unchanged without structured educational programs [Perveen et al. (14); Gad Ahmed et al. (15); Hussein et al. (19); Abo Hamed (23); Samia et al. (25); Sayed et al. (28)], supporting the conclusion that observed benefits in previous research are primarily due to the educational intervention rather than extraneous influences.

The intervention group showed highly significant changes across all knowledge domains following the implementation of the educational program ($p<0.001$). The most substantial gain was observed in overall knowledge, with a mean difference of 15.90 (95% CI: 14.32–17.49). Knowledge of causes also showed a marked improvement (mean difference: 7.77), followed by knowledge of signs and symptoms (mean difference: 3.80), difficulties (mean difference: 3.37), treatment (mean difference: 2.55), and general knowledge (mean difference: 1.58). These results illustrate that the training program significantly improved mothers' knowledge of CP. The post-intervention results of the this study demonstrated a statistically significant improvement in the intervention group's knowledge, while no significant change was observed in the control group. These findings are consistent with previous studies [(Rafique (5); Perveen et al. (14); Gad Ahmed et al. (15); Samia et al. (25)] reporting that structured educational programs considerably improve caregivers' knowledge in a variety of domains.

Improvements in caregiver knowledge were observed, increasing from 12% pre-intervention to 94% post-intervention; however, these gains declined over time, underscoring the need for ongoing reinforcement. Persistent misconceptions about CP management remain problematic, indicating that single-session interventions may not ensure lasting behavior change. Nevertheless, the study provides evidence that structured educational programs can effectively improve caregiving practices, with outcomes influenced by caregivers' educational background, motivation, and resource availability, as evidenced by the Assessment of Caregivers' Knowledge. Previous studies by Gad Ahmed et al. (15) and Abo Hamed (23) found that educational and training programs can successfully improve caregiver knowledge. However, both studies emphasized that without ongoing reinforcement, the advantages may diminish, and one-time interventions are frequently ineffective. They stated that caregiver motivation, education, and available resources substantially influence the effective application of new knowledge to daily care practices.

Study Limitations

This study has certain limitations. The short follow-up duration prevented evaluation of long-term knowledge retention. Variability in caregivers' educational backgrounds and learning abilities may have influenced the outcomes of the intervention. Moreover, caregivers' household responsibilities and difficulties with transport occasionally restricted their full participation in the training sessions. Finally, limited availability of rehabilitation services in the region may have affected caregivers' ability to consistently apply the learned home-care skills.

Conclusion

The study reveals that the caregivers of children with CP are predominantly mothers in their thirties. It underscores differences in socioeconomic status, job availability, and literacy, which affect the quality of care and access to resources. Children with CP exhibit diverse characteristics, including age, gender, and CP type. The study recommended the following:

- Home care visits to improve caregivers' performance in CP home care.
- Focused educational programs to enhance caregivers' knowledge of CP.
- Ongoing and personalized educational programs to address varied caregiver backgrounds and ensure lasting improvements.
- Further research with longer follow-up periods is recommended to evaluate knowledge retention and the long-term impact on child outcomes.

Ethics

Ethics Committee Approval: The study was approved by the Ethical Committee of the Hawler Medical University, College of Nursing (approval number: 2425, date: 22.08.2024).

Informed Consent: Verbal informed consent was obtained from all caregivers who participated in the study.

Footnotes

Authorship Contributions: Concept - S.S.H.; Design - S.S.H.; Data Collection or Processing - H.F.M., Analysis or Interpretation - H.F.M., S.S.H.; Literature Search - S.S.H.; Writing - S.S.H.

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. Almosallam A, Qureshi AZ, Alzahrani B, AlSultan S, Alzubaidi WI, Alsanad A. Caregiver knowledge, attitude, and behavior toward care of children with cerebral palsy: a Saudi Arabian perspective. Healthcare (Basel). 2024; 12: 982.
2. Abd Elmagid DS, Magdy H. Evaluation of risk factors for cerebral palsy. Egypt J Neurol Psychiatry Neurosurg. 2021; 57: 13.
3. Mohamed O. An electronic game program to improve some executive functions in children with cerebral palsy. Rsch. 2021; 3: 636-715.
4. Lewis SA, Ruttenberg A, Iyiyol T, Kong N, Jin SC, Kruer MC. Potential clinical applications of advanced genomic analysis in cerebral palsy. EBioMedicine. 2024; 106: 105229.
5. Rafique A. Negative air ions induced amelioration of biochemical parameters in cerebral palsy patients. Liaquat Med Res J. 2024; 6: 10-8.
6. Mushta SM, Alghamdi R, Almalki H, Waqas S, Alawwadh A, Barasheed O, et al. Saudi Cerebral Palsy Register (SCPR): protocol on the methods and technical details. J Epidemiol Glob Health. 2024; 14: 453-61.
7. Yaqob KS, Karim SK. Children and cerebral palsy: the hopeful clinical and neuro-imaging impacts of joint therapy. EC Nutrition. 2017; 11: 176-82.
8. Asiri FYI, Tennant M, Kruger E. Oral health of individuals with cerebral palsy in Saudi Arabia: a systematic review. Community Dent Oral Epidemiol. 2024; 52: 121-9.
9. Nikolovska L, Ismani B. Effectiveness of physical therapy and rehabilitation in improving motor skills and mental function in children with cerebral palsy. MEDIS Int J Med Sci Res. 2023; 2(4): 23-6.
10. Evensen TL, Vik T, Andersen GL, Bjellmo S, Hollung SJ. Prevalence, birth, and clinical characteristics of dyskinetic cerebral palsy compared with spastic cerebral palsy subtypes: a Norwegian register-based study. Dev Med Child Neurol. 2023; 65: 1464-74.
11. Al-Jabri BA, Al-Amri AS, Jawhari AA, Sait RM, Talb RY. Prevalence, types, and outcomes of cerebral palsy at a tertiary center in Jeddah, Saudi Arabia. Cureus. 2022; 14: e27986.
12. Rashad LT, El-Dakhakhny AM, Abd Elsalam EA, Mohamed BM. Effectiveness of maternal training program on improvement of care provided to their children with cerebral palsy at Zagazig University Hospitals. Zagazig Nurs J. 2021; 17: 1-12.
13. Poojari DP, Umakanth S, Maiya GA, et al. Effect of family-centered care interventions on well-being of caregivers of children with cerebral palsy: a systematic review. F1000Res. 2024; 12: 790.
14. Perveen S, Ali A, Sarwar H. Knowledge and care among mothers having children with cerebral palsy. The Therapist. 2023; 4: 25-30.
15. Gad Ahmed S, Khalil A, Abou Zeid M. Effect of an educational program about cerebral palsy management on mothers' performance. Port Said Sci J Nurs. 2023; 10: 59-79.
16. Jensen A. Cerebral palsy - brain repair with stem cells. J Perinat Med. 2022; 51: 737-51.
17. Salim Faqi-Rahid F, Hasan S. Caregiver's knowledge regarding homecare management of spastic cerebral palsy children under 12 years old in Rapareen Province. Erbil J Nurs Midwifery. 2022; 5.
18. Ajayi OO, Fatudimu MB, Hamzat TK. Effects of an educational programme on knowledge of cerebral palsy among informal caregivers of children presenting at selected physiotherapy clinics in Ibadan. Eur J Med Health Sci. 2022; 4: 57-62.
19. Hussein H, El Awady S, El Afandy A. Mothers' perception toward their children who suffering from cerebral palsy at the pediatric outpatient in Minia University Hospital. J Biosci Appl Res. 2023; 9: 292-310.
20. Eloreidi A, Kehyayan V. Needs of caregivers of children with cerebral palsy: a literature review. J Nurs Educ Pract. 2021; 11: 12-9.
21. Moenardi MCS, Sungkar E, Hawani D. Cerebral palsy parents' knowledge, attitude, and behavior at Dr. Hasan Sadikin General Hospital Bandung 2014. Althea Med J. 2020; 7: 84-8.
22. Bibi S, Ashfaq M, Dars JA, Ahmed SH, Ahmed A. Stress level in parents of children with cerebral palsy. Prof Med J. 2024; 31: 862-7.
23. Abo Hamed, N. Assessment caregiver's knowledge and practice regarding care of their children with cerebral palsy. Helwan International Journal for Nursing Research and Practice. 2023; 2: 206-18.
24. Ramadan Mohamed Ebeed S, El-Sayed Isamail Omar T, El-Sayed Youssef M, Ibrahim Mohamed Hamad N. Quality of life of mothers having children with cerebral palsy. Alex Sci Nurs J. 2024; 26: 15-27.
25. Samia P, Tirkha M, Kassam AI, Muindi R, Gitaka W, Wamithi S, et al. Quality of life in a cohort of Kenyan children with cerebral palsy. Glob Pediatr. 2024; 9: 100172.
26. Gamal Mohammed Desoky D, Abd El-Hakeem Ahmed H, Fouad Malika F. Quality of life and its effect on mothers providing care for children with cerebral palsy. Egypt J Health Care. 2021; 12: 1369-86.
27. Jahan I, Sultana R, Muhit M, Akbar D, Karim T, Al Imam MH, et al. Nutrition interventions for children with cerebral palsy in low- and middle-income countries: a scoping review. Nutrients. 2022; 14: 1211.
28. Sayed SA, Hammad EM, Sharkawy S, Abdalfata W. Mothers' knowledge, attitudes, and practices towards children with cerebral palsy at Assiut City, Egypt. Assiut Sci Nurs J. 2024; 12: 169-83.

İstanbul Med J 2026; 27(2): 171-6
DOI: 10.4274/imj.galenos.2026.27981

Comparison of Macular Choroidal Thickness in Treatment-Naive Normal-Tension Glaucoma and Ocular Hypertension Patients

Furkan Çiftci[1], Sibel Zırtıloğlu[2], Mustafa Suat Alıkma[3]

[1]University of Health Sciences Türkiye, Mersin City Hospital, Clinic of Ophthalmology, Mersin, Türkiye
[2]University of Health Sciences Türkiye, Kanuni Sultan Suleyman Training and Research Hospital, Clinic of Ophthalmology, İstanbul, Türkiye
[3]Servergazi State Hospital, Clinic of Ophthalmology, Denizli, Türkiye

ABSTRACT

Introduction: To evaluate foveal, temporal parafoveal, and nasal parafoveal choroidal thicknesses together with retinal nerve fiber layer (RNFL) parameters in treatment-naive patients with normal-tension glaucoma (NTG), and to compare these findings with those of patients with ocular hypertension (OHT).

Methods: This retrospective cross-sectional study included one eye from each of 40 NTG patients and 66 OHT patients. In all participants, quadrant- and clock-hour-based RNFL measurements and foveal, temporal-parafoveal, and nasal-parafoveal choroidal thicknesses were assessed using optical coherence tomography.

Results: No significant differences were observed between the groups regarding age, sex, axial length, or disc area. Mean RNFL thickness and quadrant-based and clock-hour RNFL values were significantly lower in the NTG group than in the OHT group at all locations except the 9 o'clock sector ($p<0.05$). While no significant intergroup differences were detected in foveal or nasal parafoveal choroidal thickness, temporal parafoveal choroidal thickness was greater in the OHT group than in the NTG group ($p=0.047$).

Conclusion: Based on the present results, macular choroidal thickness by itself does not appear to provide sufficient discriminatory value in normal-tension glaucoma. Rather than relying solely on macular choroidal measurements, microvascular changes in the optic nerve head may represent a more relevant contributor to glaucomatous damage. Future investigations integrating functional vascular imaging modalities may clarify the mechanisms underlying the pathophysiology of normal-tension glaucoma.

Keywords: Normal-tension glaucoma, ocular hypertension, choroidal thickness, macular choroidal thickness

Introduction

Glaucoma is characterized by progressive loss of retinal ganglion cells, leading to irreversible visual field (VF) impairment over time (1). Although aging and elevated intraocular pressure (IOP) are considered the principal risk factors for glaucoma development, the presence of progressive optic nerve damage in some patients despite normal IOP levels suggests that pressure-independent mechanisms also contribute to disease pathogenesis (2-4). Among these mechanisms, neurodegenerative processes have been proposed to play a role in normal-tension glaucoma (NTG), whereby mitochondrial dysfunction in retinal ganglion cells, reduced energy metabolism, and increased oxidative stress may increase cellular vulnerability (5).

In patients with NTG, the pathophysiology of retinal ganglion cell loss is frequently associated with impaired microcirculation at the optic nerve head (ONH) and insufficient peripapillary choroidal perfusion (3). While the inner retinal layers are supplied by the retinal vasculature, the metabolic demands of the retinal pigment epithelium and photoreceptor layers are largely met by the choroidal circulation (6,7). Accordingly, choroidal structures are thought to play a role in glaucoma pathogenesis, particularly with respect to vascular mechanisms.

Variability in choroidal thickness has been attributed to multiple ocular and systemic determinants, such as age-related changes and axial length (AL) differences, which complicates its interpretation in glaucomatous eyes (8,9). Previous studies evaluating the relationship between choroidal thickness and glaucomatous damage have reported inconsistent findings. Some studies reported an association between choroidal parameters and glaucomatous injury (10-12). However, others found no significant differences between glaucoma patients and healthy individuals (9,13,14). Furthermore, studies assessing choroidal structures in OHT and glaucoma

Address for Correspondence: Furkan Çiftci, MD, University of Health Sciences Türkiye, Mersin City Hospital, Clinic of Ophthalmology, Mersin, Türkiye
E-mail: drfurkanciftci@gmail.com **ORCID ID:** orcid.org/0000-0002-9415-5630

Cite this article as: Çiftci F, Zırtıloğlu S, Alıkma MS. Comparison of macular choroidal thickness in treatment-naive normal-tension glaucoma and ocular hypertension patients. İstanbul Med J. 2026; 27(2): 171-6

Received: 18.03.2026
Accepted: 27.04.2026
Publication Date: 12.05.2026

have indicated that choroidal alterations do not occur uniformly across all glaucoma subtypes or anatomical regions (15-17).

In the present study, foveal, temporal parafoveal, and nasal parafoveal choroidal thicknesses, together with quadrant- and clock-hour–based retinal nerve fiber layer (RNFL) parameters, were evaluated in treatment-naive patients with NTG and compared with those of patients with OHT who showed no RNFL damage or VF abnormalities and had a cup-to-disc (C/D) ratio ≤0.3. OHT patients were selected as the control group instead of healthy individuals to permit a more specific assessment of choroidal changes potentially related to glaucomatous damage in eyes with comparable optic disc morphology and without structural glaucomatous injury. The primary aim of this study was to evaluate macular choroidal thickness in treatment-naive NTG patients and to compare these findings with those of OHT patients.

Methods

Study Population

This retrospective cross-sectional study was conducted in accordance with the Declaration of Helsinki and was approved by the Pamukkale University Ethics Committee (approval number: 60116787-020/956, date: 15.01.2020). Ophthalmological examination findings and OCT data from treatment-naive patients with NTG and OHT who presented to the clinic between March 2017 and January 2020 were reviewed retrospectively. Because of the retrospective design of the study, the requirement for informed consent was waived by the ethics committee.

Inclusion criteria for the NTG group were defined as a newly established diagnosis of NTG, glaucomatous optic neuropathy characterized by neuroretinal rim thinning, and detectable RNFL defects. In addition, at least three reliable VF tests were required, with a minimum of two fulfilling the Anderson criteria (fixation losses <20% and false-positive/false-negative rates <10%). IOP was required to remain consistently below 21 mmHg, based on measurements obtained on three separate days at different times. Central corneal thickness was required to be within the range of 555–588 µm, and patients with prior glaucoma treatment were excluded.

Eligibility criteria for the OHT group included an IOP greater than 21 mmHg, a C/D ratio ≤0.3 in the absence of glaucomatous optic disc changes, and a central corneal thickness between 555 and 588 µm. Additional requirements were the absence of RNFL defects on OCT, normal findings on standard automated perimetry, and the absence of previous glaucoma treatment.

Only one eye per participant was included in the analysis. When both eyes met the eligibility criteria, the right eye was selected for analysis. Gonioscopy was performed in all cases to confirm an open anterior chamber angle and to exclude additional pathology.

Eyes with best-corrected visual acuity worse than 20/20 or spherical refractive error outside the range of −6.00 to +6.00 diopters were excluded. Other exclusion criteria included a history of ocular surgery or trauma, prior glaucoma treatment, systemic diseases (such as diabetes mellitus, hypertension, and inflammatory vascular disorders), amblyopia, obstructive sleep apnea syndrome, and obesity.

Demographic characteristics (age and sex) and findings from a comprehensive ophthalmologic examination were recorded for all participants. The analysis included VF indices, IOP, AL, and OCT-derived measurements.

Measurements

IOP measurements were obtained by means of Goldmann applanation tonometry, and AL was obtained by A-scan ultrasonography. Mean, quadrant-based, and clock-hour RNFL thicknesses, rim area, disc area, mean and vertical C/D ratios, and cup volume were assessed using the Optic Disc Cube 200 × 200 OCT protocol.

Foveal, temporal parafoveal, and nasal parafoveal choroidal thicknesses were measured using the 5-Line Raster OCT protocol. Foveal choroidal thickness was operationally defined as the vertical distance extending from the posterior margin of the retinal pigment epithelium to the choroid–sclera junction. Parafoveal measurements in the temporal and nasal regions were obtained at points located 500 µm from the foveal center on either side, using the same method. Choroidal thickness measurements were performed manually using the caliper tool provided by the OCT software.

OCT assessments were performed using a Cirrus HD-OCT platform (Carl Zeiss Meditec, Dublin, CA, USA), and images were acquired following standardized scanning procedures. All OCT measurements were performed by a single examiner using the same device and standard scanning procedures. The individual was unaware of the subjects' group assignments. The following parameters were recorded and statistically compared between the NTG and OHT groups: IOP, AL, RNFL thicknesses (mean, quadrant-based, and clock-hour), rim area, disc area, mean and vertical C/D ratios, cup volume, and foveal, temporal parafoveal, and nasal parafoveal choroidal thicknesses.

Statistical Analysis

Statistical analyses were performed using the SPSS software package (version 25.0, SPSS Inc., Chicago, IL, USA). Quantitative variables are reported as mean ± standard deviation together with minimum and maximum values, while categorical data are summarized as frequencies and percentages. The normality of the distribution was evaluated using the Kolmogorov–Smirnov test. For normally distributed parameters, intergroup comparisons were performed using the independent-samples t-test. Statistical significance was defined as a two-sided p value less than 0.05. Correction for multiple testing was not applied. No a priori sample size calculation was performed due to the retrospective design. All eligible patients within the study period were included.

Results

A total of 66 eyes from 66 patients with OHT and 40 eyes from 40 patients with NTG were included in the study. Demographic characteristics and baseline ocular parameters were comparable between groups, and no significant differences were observed in age, AL, or disc area (Table 1).

Choroidal thickness measurements are presented in Table 2. Temporal parafoveal choroidal thickness was significantly greater in the OHT group compared with the NTG group, whereas no significant intergroup

differences were observed in foveal or nasal parafoveal choroidal thickness.

In contrast, several ONH and RNFL parameters differed significantly between the groups. The NTG group demonstrated significantly lower mean RNFL thickness, as well as reduced RNFL thicknesses in the superior, temporal, inferior, and nasal sectors compared with the OHT group. In addition, rim area was significantly smaller in the NTG group, whereas mean and vertical C/D ratios and cup volume were significantly greater (Table 3, Figure 1).

Clock-hour analysis revealed significantly lower RNFL thickness in most sectors in the NTG group compared with the OHT group. The only sector without a significant intergroup difference was the 9 o'clock position (Table 3).

Table 1. Demographic and baseline ocular characteristics of the ocular hypertension and normal-tension glaucoma groups

Parameter	Ocular hypertension (group 1) n=66	Normotensive glaucoma (group 2) n=40	p value
Age (years)	51–69 (59.5±4.4)	51–69 (60.2±5.1)	0.480
Sex (M/F)	20/46 (30.3/69.7)	18/22 (45.0/55.0)	0.187
Axial length (mm)	21.86–25.32 (23.24±0.67)	22.05–25.35 (23.28±0.68)	0.763
Disc area (mm²)	1.32–3.19 (1.99±0.30)	1.06–2.89 (1.98±0.42)	0.904

*Data are presented as minimum–maximum (mean ± standard deviation). Between-group comparisons were performed using the independent samples t-test. A p value <0.05 was considered statistically significant. M/F: Male/female

Table 2. Comparison of macular choroidal thickness parameters between the ocular hypertension and normal-tension glaucoma groups

Parameter	Ocular hypertension (group 1) n=66	Normotensive glaucoma (group 2) n=40	p value
Temporal parafoveal choroidal thickness (µm)	150–296 (213.09±31.7)	131–278 (199.6±36.3)	0.047
Foveal choroidal thickness (µm)	175–316 (243.76±31)	168–304 (236.6±34.1)	0.270
Nasal parafoveal choroidal thickness (µm)	160–298 (222.89±33)	146–280 (216.9±35.9)	0.389

Table 3. Comparison of intraocular pressure, optic nerve head, and RNFL parameters between ocular hypertension and normal-tension glaucoma groups

Parameter	Ocular hypertension (Group 1) n=66	Normal-tension glaucoma (Group 2) n=40	p value
Intraocular pressure (mmHg)	24–28 (24.5±1.3)	13–20 (15.75±1.7)	<0.001
Average RNFL thickness (µm)	78–110 (94.47±6.8)	55–96 (74.3±9)	<0.001
Rim area (mm²)	0.92–1.97 (1.35±0.2)	0.37–1.6 (0.89±0.22)	<0.001
Average C/D ratio	0.3–0.77 (0.53±0.11)	0.28–0.84 (0.71±0.1)	<0.001
Vertical C/D ratio	0.15–0.73 (0.49±0.11)	0.47–0.86 (0.7±0.09)	<0.001
Cup volume (mm³)	0.004–0.771 (0.198±0.166)	0.003–0.812 (0.421±0.227)	<0.001
Superior RNFL thickness (µm)	93–150 (118.7±11)	64–130 (86.6±17.2)	<0.001
Temporal RNFL thickness (µm)	52–98 (68.6±10.6)	42–83 (55.9±10.5)	<0.001
Inferior RNFL thickness (µm)	55–144 (119.5±15.6)	50–132 (89.8±19.4)	<0.001
Nasal RNFL thickness (µm)	23–113 (69.98±10.5)	34–93 (62.6±12.6)	0.002
RNFL thickness clock hour 1 (µm)	44–173 (106.1±23.4)	44–150 (79.6±24.7)	<0.001
RNFL thickness clock hour 2 (µm)	59–127 (81.5±15.3)	37–99 (66.5±14.2)	<0.001
RNFL thickness clock hour 3 (µm)	39–84 (56.4±12.3)	31–72 (51.3±10.2)	0.033
RNFL thickness clock hour 4 (µm)	48–98 (67.8±12.5)	32–76 (57.1±10.1)	<0.001
RNFL thickness clock hour 5 (µm)	68–188 (116.7±27.4)	39–162 (86.2±24.6)	<0.001
RNFL thickness clock hour 6 (µm)	79–171 (128.9±23)	58–147 (100.6±24.9)	<0.001
RNFL thickness clock hour 7 (µm)	55–177 (117.8±30.9)	41–142 (90.2±23.0)	<0.001
RNFL thickness clock hour 8 (µm)	45–107 (69.8±12.9)	36–89 (57.6±12.5)	<0.001
RNFL thickness clock hour 9 (µm)	32–84 (55±9.4)	34–86 (54.7±12.3)	0.855
RNFL thickness clock hour 10 (µm)	56–146 (82±16.3)	46–103 (72.7±17.5)	0.007
RNFL thickness clock hour 11 (µm)	85–173 (123.3±21.1)	43–149 (90±28.2)	<0.001
RNFL thickness clock hour 12 (µm)	80–168 (125.3±22.2)	59–162 (91.6±26.8)	<0.001

RNFL: Retinal nerve fiber layer, C/D: Cup-to-disc

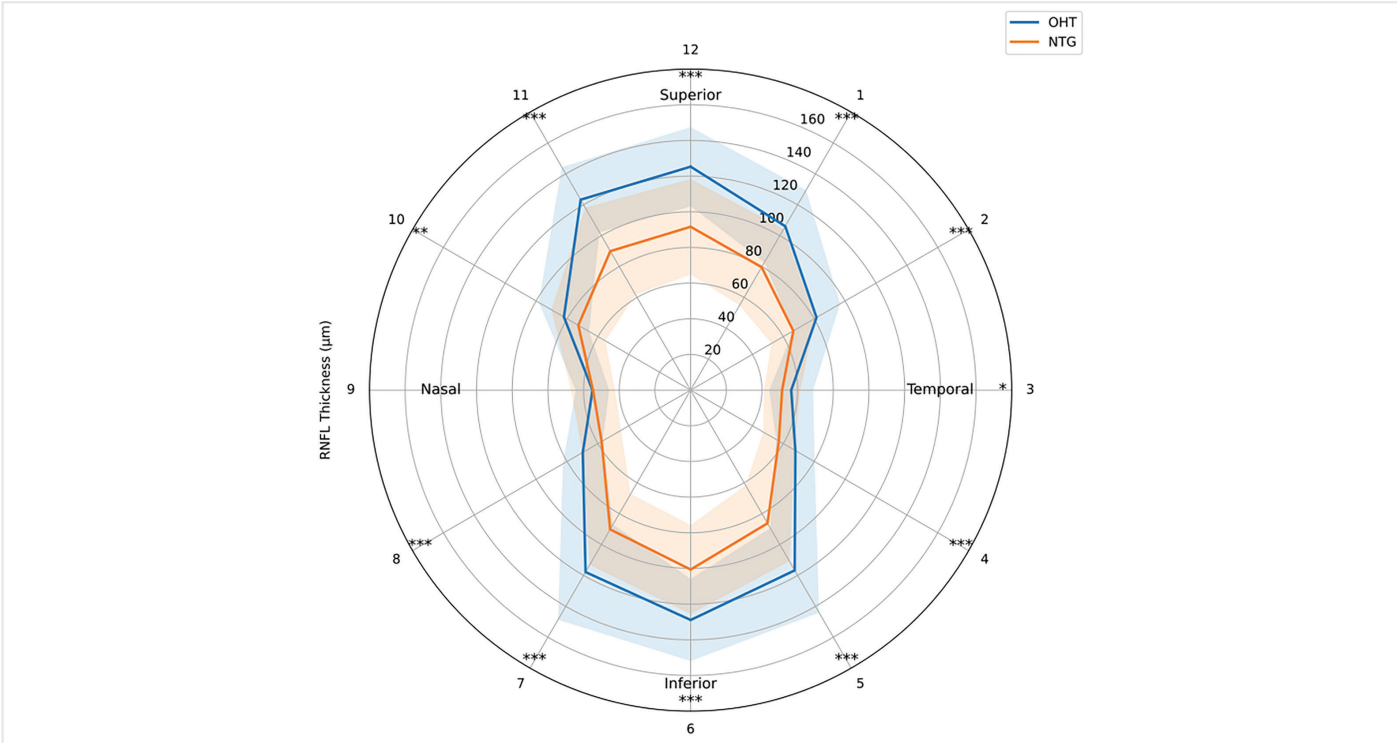

Figure 1. Clock-hour distribution of peripapillary retinal nerve fiber layer (RNFL) thickness in the ocular hypertension and normal-tension glaucoma (NTG) groups. RNFL thickness was generally lower in the NTG group across most sectors, with no significant difference in the 9 o'clock sector. Values are presented as mean ± SD. Statistical significance is indicated as *p<0.05, **p<0.01, ***p<0.001
SD: Standard deviation, OHT: Ocular hypertension

Discussion

Although VF loss detected by perimetry remains central to the diagnosis and monitoring of glaucoma, it should be interpreted in conjunction with RNFL loss and structural alterations of the ONH. OCT derived RNFL and ONH parameters provide a quantitative and objective assessment of the structural components of glaucomatous damage (18). Nevertheless, pressure-independent vascular factors have also been reported to contribute to disease pathogenesis, particularly in NTG (3). Given their roles in retinal metabolism and peripapillary circulation, choroidal structures have been considered potential contributors to glaucomatous damage (12,19,20). Accordingly, in the present study, foveal, temporal parafoveal, and nasal parafoveal choroidal thicknesses were evaluated, together with quadrant- and clock-hour–based RNFL parameters, in treatment-naive NTG patients and compared with those in individuals with OHT.

Reduced RNFL thickness is closely associated with glaucoma, and mean, quadrant-based, and clock-hour RNFL measurements, together with rim area and vertical C/D ratio, are regarded as key structural indicators of glaucomatous damage (21-25). RNFL thinning particularly in the 6th, 7th, and 8th clock-hour sectors has been reported to be strongly associated with glaucomatous injury (26). Previous studies in NTG have demonstrated significant reductions in RNFL thickness in the mean, superior, and inferior quadrants (27). Most published studies have compared patients with glaucoma to healthy controls. In contrast, the present study compared NTG patients with OHT patients without structural or functional glaucomatous damage. Significant thinning was observed in the NTG group, not only in mean RNFL thickness but also in all four quadrants (superior, temporal, inferior, and nasal). Clock-hour analysis further demonstrated a reduction in RNFL thickness in all sectors except at the 9 o'clock position compared with the OHT group, indicating the coexistence of diffuse and regional nerve fiber loss in NTG.

Choroidal thickness is influenced by multiple factors, including age, AL, other ocular biometric parameters, and systemic conditions (12,13). In the present study, choroidal thickness was evaluated in treatment-naive patients diagnosed with NTG and OHT. Demographic characteristics such as age, sex, and AL were comparable between the two groups. To minimize the potential impact of glaucoma medications on choroidal measurements, only untreated individuals were included.

While some studies have reported that glaucoma therapy may be associated with increased choroidal thickness (14), others have found no significant difference between glaucomatous eyes and healthy controls in this regard (22,28). Similarly, a study investigating unilateral glaucoma reported no meaningful interocular difference in choroidal thickness (29). Taken together, these findings suggest that choroidal structural changes are not uniform across all types of glaucoma and may vary depending on the anatomical region examined. In line with these observations, the present study did not demonstrate a significant difference in foveal and nasal-parafoveal choroidal thicknesses between NTG and OHT patients, although a regional difference was observed in the temporal parafoveal area.

Contradictory findings regarding choroidal thickness in patients with OHT and NTG have been reported in the literature. In OHT, elevated

IOP may lead to choroidal thinning through mechanical compression. In contrast, in NTG, vascular dysfunction and microcirculatory disturbances may affect choroidal structure despite normal IOP levels. The coexistence of these mechanisms may partly account for the difficulty in identifying clear differences in choroidal thickness between the two groups.

Previous studies in OHT patients have suggested an association between choroidal thinning and impaired ocular hemodynamics. Bayraktar et al. (20) reported that reduced choroidal thickness may be related to increased vascular resistance and decreased ocular blood flow parameters. Similarly, Yilmaz et al. (30) demonstrated reduced macular choroidal thickness, particularly in the temporal region, compared with healthy controls. These differences may be related to variations in patient characteristics, disease severity, or measurement methods.

Choroidal thickness is a structural parameter and does not directly reflect choroidal perfusion. In the present study, temporal parafoveal choroidal thickness was higher in the OHT group than in the NTG group ($p=0.047$). However, this finding should be interpreted with caution because it was limited to a single macular region, was of borderline statistical significance, and was not accompanied by consistent differences in the foveal or nasal parafoveal areas. This isolated result may reflect regional variability, measurement-related factors associated with manual caliper assessment, or limited sample size, rather than a stable disease-specific pattern. In addition, discrepancies with previous studies may be related to differences in treatment status, patient characteristics, anatomical regions analyzed, and measurement methodology. Taken together, these findings further support the view that macular choroidal thickness alone has limited discriminatory value in differentiating NTG from OHT.

Importantly, the absence of a significant difference in choroidal thickness should not be interpreted as evidence against a role for choroidal circulation in glaucomatous damage. Microvascular alterations at the level of the ONH may not be adequately reflected by macular choroidal thickness measurements. Therefore, functional vascular assessments may provide more meaningful insights than structural measurements in understanding the pathophysiology of NTG, highlighting the need for further studies supported by advanced vascular imaging techniques.

Study Limitations

This study has several limitations. The retrospective design does not allow for causal inferences. Interobserver and intraobserver reproducibility were not formally assessed. In addition, the relatively small sample size may have reduced the sensitivity of some analyses.

Finally, peripapillary choroidal thickness was not evaluated. Given the potential role of the ONH microenvironment in glaucomatous damage, this aspect may not have been fully addressed in the present study.

Conclusion

In this study, foveal and nasal parafoveal choroidal thicknesses were comparable between patients with NTG and OHT, whereas only a marginal difference was observed in the temporal parafoveal region. Given the number of regional comparisons performed, this finding should be interpreted with caution. Overall, the results indicate that macular choroidal thickness alone may not be sufficient to distinguish normal-tension glaucoma from ocular hypertension. It is possible that microvascular changes at the level of the ONH play a more relevant role in disease mechanisms than macular choroidal structure alone. Further studies incorporating functional vascular imaging techniques may provide a more comprehensive understanding of the underlying pathophysiology of NTG.

Ethics

Ethics Committee Approval: This retrospective cross-sectional study was conducted in accordance with the Declaration of Helsinki and was approved by the Pamukkale University Ethics Committee (approval number: 60116787-020/956, date: 15.01.2020).

Informed Consent: Because of the retrospective design of the study, the requirement for informed consent was waived by the ethics committee.

Footnotes

Authorship Contributions: Surgical and Medical Practices - M.S.A.; Concept - M.S.A.; Design - F.Ç., S.Z., M.S.A.; Data Collection or Processing - M.S.A.; Analysis or Interpretation - F.Ç., S.Z.; Literature Search - F.Ç., S.Z.; Writing - M.S.A.

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. Yabana T, Sato K, Shiga Y, Himori N, Omodaka K, Nakazawa T. The relationship between glutathione levels in leukocytes and ocular clinical parameters in glaucoma. PLoS One. 2019; 14: e0227078.
2. Coleman AL, Miglior S. Risk factors for glaucoma onset and progression. Surv Ophthalmol. 2008; 53: S3-10.
3. Trivli A, Koliarakis I, Terzidou C, Goulielmos GN, Siganos CS, Spandidos DA, et al. Normal-tension glaucoma: pathogenesis and genetics. Exp Ther Med. 2019; 17: 563-74.
4. Flammer J, Orgül S, Costa VP, Orzalesi N, Krieglstein GK, Serra LM, et al. The impact of ocular blood flow in glaucoma. Prog Retin Eye Res. 2002; 21: 359-93.
5. Osborne NN. Pathogenesis of ganglion "cell death" in glaucoma and neuroprotection: focus on ganglion cell axonal mitochondria. Prog Brain Res. 2008; 173: 339-52.
6. Funk RH. Blood supply of the retina. Ophthalmic Res. 1997; 29: 320-5.
7. Sezer T, Altınışık M, Koytak İA, Özdemir MH. Choroid and optical coherence tomography. Turk J Ophthalmol. 2016; 46: 30-7.
8. Zhou H, Dai Y, Shi Y, Russell JF, Lyu C, Noorikolouri J, et al. Age-related changes in choroidal thickness and the volume of vessels and stroma using swept-source oct and fully automated algorithms. Ophthalmol Retina. 2020; 4: 204-15.
9. Maul EA, Friedman DS, Chang DS, Boland MV, Ramulu PY, Jampel HD, et al. Choroidal thickness measured by spectral domain optical coherence tomography: factors affecting thickness in glaucoma patients. Ophthalmology. 2011; 118: 1571-9.
10. Bayraktar S, Cebeci Z, Izgi B, Kasali K. The effect of glaucoma medication on choroidal thickness measured with enhanced depth-imaging optical coherence tomography. Med Hypothesis Discov Innov Ophthalmol. 2019; 8: 44-51.

11. Roberts KF, Artes PH, O'Leary N, Reis AS, Sharpe GP, Hutchison DM, et al. Peripapillary choroidal thickness in healthy controls and patients with focal, diffuse, and sclerotic glaucomatous optic disc damage. Arch Ophthalmol. 2012; 130: 980-6.
12. Hirooka K, Fujiwara A, Shiragami C, Baba T, Shiraga F. Relationship between progression of visual field damage and choroidal thickness in eyes with normal-tension glaucoma. Clin Exp Ophthalmol. 2012; 40: 576-82.
13. Hirooka K, Tenkumo K, Fujiwara A, Baba T, Sato S, Shiraga F. Evaluation of peripapillary choroidal thickness in patients with normal-tension glaucoma. BMC Ophthalmol. 2012;12:29.
14. Nakakura S, Yamamoto M, Terao E, Nagasawa T, Tabuchi H, Kiuchi Y. The whole macular choroidal thickness in subjects with primary open angle glaucoma. PLoS One. 2014; 9: e110265.
15. Akil H, Al-Sheikh M, Falavarjani KG, Francis B, Chopra V. Choroidal thickness and structural glaucoma parameters in glaucomatous, preperimetric glaucomatous, and healthy eyes using swept-source OCT. Eur J Ophthalmol. 2017; 27: 548-54.
16. Lamparter J, Schulze A, Riedel J, Wasielica-Poslednik J, König J, Pfeiffer N, et al. Peripapillary choroidal thickness and choroidal area in glaucoma, ocular hypertension and healthy subjects by SD-OCT. Klin Monbl Augenheilkd. 2015; 232: 390-4.
17. Hosseini H, Nilforushan N, Moghimi S, Bitrian E, Riddle J, Lee GY, et al. Peripapillary and macular choroidal thickness in glaucoma. J Ophthalmic Vis Res. 2014; 9: 154-61.
18. Dagdelen K, Dirican E. The assessment of structural changes on optic nerve head and macula in primary open angle glaucoma and ocular hypertension. Int J Ophthalmol. 2018; 11: 1631-7.
19. Fan N, Wang P, Tang L, Liu X. Ocular blood flow and normal tension glaucoma. Biomed Res Int. 2015; 2015: 308505.
20. Bayraktar S, İpek A, Takmaz T, Yildiz Tasci Y, Gezer MC. Ocular blood flow and choroidal thickness in ocular hypertension. Int Ophthalmol. 2022; 42: 1357-68.
21. Na JH, Sung KR, Baek S, Sun JH, Lee Y. Macular and retinal nerve fiber layer thickness: which is more helpful in the diagnosis of glaucoma? Invest Ophthalmol Vis Sci. 2011; 52: 8094-101.
22. Jonas JB. Glaucoma and choroidal thickness. J Ophthalmic Vis Res. 2014; 9: 151-3.
23. Hwang YH, Kim YY. Glaucoma diagnostic ability of quadrant and clock-hour neuroretinal rim assessment using cirrus HD optical coherence tomography. Invest Ophthalmol Vis Sci. 2012; 53: 2226-34.
24. El-Naby AEA, Abouelkheir HY, Al-Sharkawy HT, Mokbel TH. Correlation of retinal nerve fiber layer thickness and perimetric changes in primary open-angle glaucoma. J Egypt Ophthalmol Soc. 2018; 111: 7-14.
25. Mokbel T, Elbendary A, El-Sharkawy H, El-Desouky W. Retinal nerve fiber layer versus optic nerve head evaluation in the diagnosis of glaucoma and glaucoma suspect patients. J Egypt Ophthalmol Soc. 2013; 106: 253-8.
26. Sánchez-Galeana CA, Bowd C, Zangwill LM, Sample PA, Weinreb RN. Short-wavelength automated perimetry results are correlated with optical coherence tomography retinal nerve fiber layer thickness measurements in glaucomatous eyes. Ophthalmology. 2004; 111: 1866-72.
27. Kim DM, Hwang US, Park KH, Kim SH. Retinal nerve fiber layer thickness in the fellow eyes of normal-tension glaucoma patients with unilateral visual field defect. Am J Ophthalmol. 2005; 140: 165-6.
28. Wang W, Zhang X. Choroidal thickness and primary open-angle glaucoma: a cross-sectional study and meta-analysis. Invest Ophthalmol Vis Sci. 2014; 55: 6007-14.
29. Mwanza JC, Sayyad FE, Budenz DL. Choroidal thickness in unilateral advanced glaucoma. Invest Ophthalmol Vis Sci. 2012; 53: 6695-701.
30. Yilmaz PT, Koz OG, Yarangumeli A, Alp MN. Macular choroidal thickness in patients with ocular hypertension as assessed using enhanced depth ımaging optical coherence tomography. Beyoglu Eye J. 2017; 4: 92-6.

İstanbul Med J 2026; 27(2): 177-8
DOI: 10.4274/imj.galenos.2026.38488

Letter to the Editor: "Which is the Best Timing for Ultrasound-Guided Transversus Abdominis Plane Block During Laparoscopic Cholecystectomy: Preoperative or Postoperative?"

Faruk Çiçekci

Selçuk University Faculty of Medicine, Department of Anesthesiology and Reanimation, Konya, Türkiye

Dear Editor,

I have read with great interest the article titled "Which is The Best Timing for Ultrasound-Guided Transversus Abdominis Plane Block During Laparoscopic Cholecystectomy: Preoperative or Postoperative?" (1). While the study addresses a relevant clinical question, I note certain methodological limitations, particularly its retrospective nature.

First, standardization of the transversus abdominis plane (TAP) block technique is critical. The literature suggests varying success rates for subcostal, lateral, and posterior approaches in controlling somatic pain after laparoscopic cholecystectomy. For upper abdominal procedures, the subcostal approach is often superior due to higher dermatomal coverage (2). In a retrospective dataset, ensuring that blocks in both groups were performed using identical techniques or anatomical levels is difficult (3). Since technical inconsistencies—rather than timing—could drive the outcomes, the authors should clarify how they standardized these variations and accounted for the performing clinicians' experience levels.

Secondly, the reliability of retrospective visual analog scale (VAS) scores and analgesic consumption data is often compromised. In the postoperative group, performing the block at the end of surgery influences initial post-anesthesia care unit scores. However, the "pre-emptive analgesia" potential of a preoperative block—its ability to prevent central sensitization—should be evaluated through total 24-hour opioid consumption and the incidence of chronic pain rather than static, early-phase VAS scores (4). The absence of long-term cumulative data weakens the conclusion regarding optimal timing.

If the observed analgesic benefit is confined to the first postoperative hour, the clinical significance of this finding remains debatable. This transient effect may reflect the pharmacokinetic profile of the local anesthetic or the surgery's duration rather than a sustainable clinical advantage. Furthermore, the authors reported significantly lower intraoperative remifentanil consumption in the preoperative group. While this finding suggests a potential blunting of the surgical stress response, it remains unclear how intraoperative analgesia was titrated. Were objective measures, such as the Bispectral index or other nociception monitoring tools, utilized to standardize remifentanil administration? Clarifying whether this reduction was a direct result of the block's efficacy or was influenced by the clinician's subjective assessment of anesthetic depth would enhance the clinical relevance of these findings.

The observation that the statistically significant difference in pain scores was confined to the first postoperative hour significantly challenges the clinical relevance of the findings. Rather than demonstrating true superiority of preoperative timing, this transient effect may simply reflect the pharmacokinetic profile of the local anesthetic or the total duration of surgery. If the surgery outlasts the peak effect of a preoperative block, the perceived advantage in the immediate recovery phase becomes a matter of timing rather than a sustainable "pre-emptive" benefit. It is essential to question whether such a short-lived difference translates into meaningful clinical outcomes, such as reduced total 24-hour opioid consumption or accelerated functional recovery.

While this study offers a perspective on the timing of TAP blocks in clinical practice, its retrospective nature necessitates support from prospective, randomized controlled trials. I believe that it would be prudent for readers to consider these limitations when integrating these data into their clinical protocols. The insights from the authors and the editorial board on these matters would be valuable to clinicians aiming to optimize patient recovery.

Footnotes

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

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

Address for Correspondence: Prof. Faruk Çiçekci, MD, Selçuk University Faculty of Medicine, Department of Anesthesiology and Reanimation, Konya, Türkiye
E-mail: farukcicekci@yahoo.com **ORCID ID:** orcid.org/0000-0002-3248-0745

Cite this article as: Çiçekci F. Letter to the Editor: "Which is the best timing for ultrasound-guided transversus abdominis plane block during laparoscopic cholecystectomy: preoperative or postoperative?" İstanbul Med J. 2026; 27(2): 177-8

Received: 05.02.2026
Accepted: 30.03.2026
Publication Date: 12.05.2026

References

1. Kölükçü V, Şahin AT, Balta MG, Genç A, Koca B. Which is the best timing for ultrasound-guided transversus abdominis plane block during laparoscopic cholecystectomy: preoperative or postoperative? Istanbul Med J. 2026; 27: 44-9.
2. Hebbard P, Fujiwara Y, Shibata Y, Royse C. Ultrasound-guided transversus abdominis plane (TAP) block. Anaesth Intensive Care. 2007; 35: 616-7.
3. El-Dawlatly AA, Turkistani A, Kettner SC, Machata AM, Delvi MB, Thallaj A, et al. Ultrasound-guided transversus abdominis plane block: description of a new technique and comparison with conventional systemic analgesia during laparoscopic cholecystectomy. Br J Anaesth. 2009; 102: 763-7. Erratum in: Br J Anaesth. 2009; 103: 622.
4. Woolf CJ. Central sensitization: implications for the diagnosis and treatment of pain. Pain. 2011; 152: S2-15.