ISSN 0104-4230
ISSN 1806-9282 (On-line)

RAMB

Journal of The Brazilian Medical Association

Volume 72, Number 3
March, 2026

RAMB
Journal of The Brazilian Medical Association

Volume 72, Number 3, March, 2026

ISSN 0104-4230
ISSN 1806-9282 (On-line)

SECTIONS

EDITORIAL

e723EDIT Beyond survival: reconstructing body image and femininity after breast cancer

LETTERS TO THE EDITOR

e20250706 Impact of uterine closure techniques following cesarean section on the development of cesarean scar defects: consigning cicatrix to history?

e20251818 Methodological concerns regarding "Use of high-dose steroid therapy: addition of anakinra in the treatment of severe COVID-19"

ARTICLES

ORIGINAL ARTICLES

e20250931 The relationship between in-hospital mortality and Naples prognostic score in patients undergoing tricuspid valve surgery

e20251229 Efficacy of a virtual multimodal exercise program on postural awareness and digital eye strain in university students: a pilot study

e20251434 Comparison of different artificial intelligence tools' answers to questions related to early intervention: ChatGPT versus Gemini

e20251460 The effect of sodium-glucose cotransporter-2 inhibitors on haemoglobin and haematocrit in diabetic kidney disease

e20251462 Whole-body magnetic resonance imaging vs. positron emission tomography-computed tomography in pediatric oncology: enhancing diagnostic precision for solid tumors

e20251467 Relationship between lower urinary system symptoms and urinary incontinence in women with gestational diabetes

e20251488 A comparative evaluation of ChatGPT and Gemini for stroke caregiver support: accuracy, safety, and response characteristics

e20251543 Tranexamic acid in arthroscopic repair of rotator cuff tears in the shoulder: a randomized trial

e20251578 The effect of 30° and 70° arthroscope use on femoral and tibial tunnel placement in arthroscopic anterior cruciate ligament reconstruction surgery

e20251605 Immersive learning in obstetrics: a 360-degree video for teaching the management of premature rupture of membranes

e20251624 Validation of the 2023 International Federation of Gynecology and Obstetrics staging system in a large retrospective cohort

e20251626 Evaluation of prognostic biomarkers for exacerbation frequency in patients with chronic obstructive pulmonary disease

e20251628 Elevated serum Raftlin levels in non-muscle invasive bladder cancer: diagnostic implications and rs690037 genotype analysis

e20251637 Evaluation of ferroptosis markers and laboratory findings in patients with juvenile idiopathic arthritis

e20251739 Examining the relationship between breastfeeding duration at age two and beyond with parenting styles and infant feeding attitudes in Turkey (2023–2024)

e20251769 Assessing the accuracy and consistency of responses from Chat Generative Pre-trained Transformer in human papillomavirus infection and protection

e20251795 Impact of maternal nutrition and gestational weight gain on perinatal outcomes in intrahepatic cholestasis of pregnancy

e20251801 Ciprofol provides safer sedation than propofol in minor gynecologic surgery: evidence from a randomized controlled trial

e20251804 Determining the relationship between breast cancer fear in women and levels of cyberchondria

e20251808 Bullying among medical residents in a large public hospital: a cross-sectional study

e20251821 Limitations of cardiac risk scores in elective non-cardiac surgeries: a Brazilian cohort

e20251838 Epidemiological profile of pregnant women undergoing first-trimester preeclampsia screening at a tertiary reference center in Brazil

e20251878 Cord insertion anomalies and adverse pregnancy outcomes: more than an incidental finding

e20251888 Evaluating chat generative pre-trained transformer responses to common patient questions on temporomandibular disorders

REVIEW ARTICLE

e20251636 Artificial intelligence to meet the psychosocial needs of cancer patients: a systematic review of randomized controlled trials

RAMB
Journal of The Brazilian Medical Association

EDITORIAL BOARD

EDITORS-IN-CHIEF

Renato Delásciо Lopes
Roseli Nomura
José Maria Soares Jr.

MANAGING EDITOR

Cesar Teixeira

ASSOCIATED EDITORS

Albert Bousso
Ana Gabriel P. Santos
Ana Pontes
Anna Andrei
Auro Del Giglio
Claudia Leite
Dani Ejzenberg
Dimas Ikeoki
Edna Frasson de S. Montero
Eduardo F. Borba
Edward Araújo Jr
Gabriel Costa Osanan
Isabel Sorpreso
Isabela Giuliano
Lilian Sadeck
Linamara Batistella
Lucia Pellanda
Paulo Kassab
Rachel Riera
Sergio C. Nahas
Werther B. W. de Carvalho

INTERNATIONAL EDITORS

Frida Leonetti
Geltrude Mingrone
Giuseppe Barbaro
Marcelo Marotti
Walter Ageno

JUNIOR EDITOR

André Zimerman

SPECIALTY EDITORS

ACUPUNCTURE

Sidney Brandão

ALLERGY AND IMMUNOLOGY

Dirceu Solé

ANAESTHESIOLOGY

Plínio da Cunha Leal

ANGIOLOGY AND VASCULAR SURGERY

Edwaldo Edner Joviliano

CARDIOLOGY

Weimar Kunz Sebba B. de Souza

CARDIOVASCULAR

Marcela da Cunha Sales

CLINICAL ONCOLOGY

Alexandre Palladino

CLINICAL PATHOLOGY / LABORATORIAL MEDICINE

André Doi

COLOPROCTOLOGY

Henrique Sarubbi Fillmann

DERMATOLOGY

Flávia Vasques Bittencourt

DIGESTIVE ENDOSCOPY

Fauze Maluf Filho

DIGESTIVE SURGERY

Fernando Antônio Siqueira Pinheiro

EMERGENCY MEDICINE

Hélio Penna Guimarães

ENDOCRINOLOGY AND METABOLISM

Paulo Augusto Carvalho de Miranda

FAMILY AND COMMUNITY MEDICINE

Leonardo Cançado Monteiro Savassi

GASTROENTEROLOGY

Frederico Passos Marinho

GENERAL SURGERY

Luiz Carlos Von Bahten

GERIATRICS AND GERONTOLOGY

Hercilio Hoepfner Junior

GYNAECOLOGY AND OBSTETRICS

Agnaldo Lopes da Silva Filho

HAND SURGERY

Antônio Tufi Neder Filho

HEAD AND NECK SURGERY

Leandro Luongo Matos

HEMATOLOGY AND HEMOTHERAPY

Fernando Ferreira Costa

HOMEOPATHY

Flavio Dantas de Oliveira

INFECTIOUS DISEASES

Alexandre Vargas Schwarzbold

INTENSIVE MEDICINE
Israel Silva Maia

INTERNAL MEDICINE
Ana Paula de Oliveira Ramos

LEGAL MEDICINE AND MEDICAL EXAMINATIONS
Rosa Amélia Andrade Dantas

MASTOLOGY
Gil Facina

MEDICAL GENETICS
Ida Vanessa D. Schwartz

NEUROSURGERY
Manoel Jacobsen Teixeira

NEPHROLOGY
Lúcio Roberto Requião Moura

NEUROLOGY
Marcondes Cavalcante França Jr.

NUCLEAR MEDICINE
Diego Pianta

NUTROLOGY
Aline Zanetta

OCCUPATIONAL MEDICINE
Andrea Franco Amoras Magalhães

OPHTHALMOLOGY
Eduardo Melani Rocha

ORTHOPAEDICS AND TRAUMATOLOGY
Sergio Luiz Checchia

OTOLARYNGOLOGY
Thiago Freire Pinto Bezerra

PAEDIATRIC
Lilian dos Santos Rodrigues Sadeck

PAEDIATRIC SURGERY
Lisieux Eyer Jesus

PATHOLOGY
Monique Freire Santana

PHYSICAL MEDICINE AND REHABILITATION
Eduardo de Melo Carvalho Rocha

PLASTIC SURGERY
Daniela Francescato Veiga

PREVENTIVE MEDICINE AND HEALTH ADMINISTRATION
Antônio Eduardo Fernandes D´Aguiar

PSYCHIATRY
Leonardo Rodrigo Baldaçara

PULMONOLOGY / PHTHISIOLOGY
Suzana Erico Tanni Minamoto

RADIOTHERAPY
Wilson José Almeida Jr.

RADIOLOGY
Alexandre Bezerra

RHEUMATOLOGY
Ricardo Machado Xavier

SPORTS MEDICINE
Neuza Mitsuanga

SURGICAL ONCOLOGY
Héber Salvador de Castro Ribeiro

TRAFFIC MEDICINE
José Heverardo da Costa Montal

THORACIC SURGERY
Juliana Dias Nascimento Ferreira

UROLOGY
Roni de Carvalho Fernandes

ASSOCIAÇÃO MÉDICA BRASILEIRA
(BRAZILIAN MEDICAL ASSOCIATION)

MANAGEMENT BOARD 2024-2026

PRESIDENT
César Eduardo Fernandes (SP)

1ST VICE-PRESIDENT
Luciana Rodrigues da Silva (BA)

2ND VICE-PRESIDENT
Nerlan Carvalho (PR)

VICE-PRESIDENTS
Etelvino Trindade (DF) – Mid-West
Claudia Navarro Lemos (MG) – Southeast
Paulo Toscano (PA) – North
Bento Bezerra Neto (PE) – Northeast
Juarez Molinari (RS) – South

GENERAL SECRETARY
Florisval Meinão (SP)

1ST SECRETARY
Maria Rita de Souza Mesquita (SP)

1ST TREASURER
Lacildes Rovella Júnior (SP)

2ND TREASURER
Fernando Tallo (SP)

ADMINISTRATIVE DIRECTOR
Akira Ishida (SP)

CULTURAL DIRECTOR
Romulo Capello (RJ)

DIRECTOR OF CORPORATE RELATIONS
Carlos Henrique Mascarenhas (MG)

DIRECTOR OF INTERNATIONAL RELATIONS
Carlos Vicente Serrano Junior (SP)

SCIENTIFIC DIRECTOR
José Eduardo Lutaif Dolci (SP)

ACADEMIC DIRECTOR
Clóvis Francisco Constantino (SP)

DIRECTOR OF MEMBER SUPPORT SERVICES
José Aurillo Rocha (CE)

DIRECTOR OF PARLIAMENTARY AFFAIRS
Luciano Gonçalves de Souza Carvalho (DF)

COMMUNICATIONS DIRECTOR
Luiz Carlos Von Bahten (PR)

RAMB - REVISTA DA ASSOCIAÇÃO MÉDICA BRASILEIRA
(JOURNAL OF THE BRAZILIAN MEDICAL ASSOCIATION)

Editors-in-Chief: Renato Delásio Lopes, José Maria Soares Jr and Roseli Nomura.
Managing Editor: César Teixeira
E-mail: ramb@amb.org.br
Website: www.ramb.org.br

ADDRESS: Rua São Carlos do Pinhal, 324
Bela Vista – São Paulo
Postal Code: 01333-903
Phone no.: (+55 11) 3178-6800 Ext. 177
The RAMB, Journal of The Brazilian Medical Association, is an official publication of the Associacao Medica Brasileira (AMB – Brazilian Medical Association), indexed in Medline, Science Citation Index Expanded, Journal Citation Reports, Index Copernicus, Lilacs, and Qualis A3 Capes databases, and licensed by Creative CommonsR.
Registered in the 1st Office of Registration of Deeds and Documents of Sao Paulo under n. 1.083, Book B, n. 2.

Publication norms are available on the website www.ramb.org.br

All rights reserved and protected by Law n. 9.610 – 2/19/1998. No part of this publication may be reproduced without prior written authorization of the AMB, whatever the means employed: electronic, mechanical, photocopying, recording or other.

THE RAMB IS INDEXED IN SCIELO - SCIENTIFIC ELECTRONIC LIBRARY ONLINE.

Editorial Production

ZEPPELINI PUBLISHERS

The advertisements and opinions published in the Ramb are the sole responsibility of the advertisers and authors.
The AMB and Zeppelini Publishers are not responsible for its content.

https://doi.org/10.1590/1806-9282.723EDIT

Beyond survival: reconstructing body image and femininity after breast cancer

Amanda Piato[1], José Roberto Morales Piato[2]*, Bruna Salani Mota[2], José Roberto Filassi[2], Edmund Chada Baracat[3], José Maria Soares Júnior[3]

Mastectomy is still indicated in specific situations, such as when the tumor is extensive or multifocal, or when the patient carries genetic mutations that increase cancer risk. Even in these cases, the contemporary perspective is different: we aim to preserve as much skin, areola, and nipple as possible, using implants or autologous flaps to reconstruct the breast in the same surgical procedure. Immediate reconstruction is preferable, as it prevents the psychological impact of sudden loss and reduces the need for future surgeries. Beyond restoring the body's contour, it helps a woman recognize herself again in the mirror, aiding her emotional and sexual recovery[1].

Not all patients, however, require mastectomy. In a significant number of cases, it is possible to perform breast-conserving surgery, removing only the tumor and a small safety margin without compromising oncologic control. For this to be feasible, diagnosis must occur in the early stages of the disease, primarily through mammography. It is in this context that oncoplastic surgery represents a turning point. By uniting the principles of oncologic surgery with the techniques of plastic surgery, it allows for the removal of the tumor with safe margins while simultaneously reshaping the breast. More than simply preserving, oncoplasty often improves aesthetic outcomes by correcting asymmetries, elevating breast position, and providing a more harmonious shape than before the surgery[1]. This approach symbolizes a paradigm shift: breast cancer treatment does not have to leave marks of mutilation. It can, with skill and sensitivity, restore beauty and bodily equilibrium, strengthening self-esteem and the sense of femininity.

Reconstruction—whether following a mastectomy or an oncoplastic breast-conserving surgery—is not merely a question of appearance. It plays a central role in the recovery of self-esteem. The woman who goes through cancer experiences various losses: of her hair, her routine, her sense of predictability, and her own self-image. Recovering the breast is also about regaining a sense of wholeness, confidence, and the power to feel alive and desirable once more[2].

Self-esteem does not mean denying the scars but reframing their meaning. They come to tell a story of overcoming, not of loss. The reconstructed woman is not a copy of who she was; she is a new version, conscious of her own strength[2,3].

Female sexuality after breast cancer is a topic that still demands space and attentive listening. Physical, hormonal, and emotional changes can interfere with desire and intimacy. However, when there is support and open dialog, the body finds new pathways to pleasure[3]. Reconstruction and oncoplastic surgery play a crucial role in this process. By restoring harmony and confidence, they also reignite the connection with one's own body—a body that was wounded but is reborn with new forms and meaning[2,3].

It is essential that oncological care includes conversations about sexuality, desire, and affection. To speak of pleasure is also to speak of life. The body image of a woman after cancer is made of scars, but also of rebirths[4]. Modern medicine understands that healing is not just about eliminating the disease, but about reconstructing meaning, identity, and a sense of belonging. Every patient who looks at herself with pride again is proof that, even after cancer, it is possible to feel whole, beautiful, and profoundly woman once more.

The path from breast cancer patient to empowered survivor is no longer defined by the disease alone, but by a modern

[1]Universidade Santo Amaro, Faculty of Medicine – São Paulo (SP), Brazil.
[2]Universidade de São Paulo, Hospital das Clínicas, Faculty of Medicine, Department of Obstetrics and Gynecology, Mastology Sector, Gynecology Discipline – São Paulo (SP), Brazil.
[3]Universidade de São Paulo, Hospital das Clínicas, Faculty of Medicine, Department of Obstetrics and Gynecology – São Paulo (SP), Brazil.
*Corresponding author: jrpiato@uol.com.br
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on February 05, 2026. Accepted on February 15, 2026.
Scientific Editor: Renato Delascio Lopes

commitment to healing that is as comprehensive as it is compassionate. Today's paradigm in surgical oncology, exemplified by oncoplastic techniques and immediate reconstruction, has elevated care beyond mere tumor eradication to actively restoring the body's harmony and a woman's self-image. This tangible restoration of form, however, is only the beginning. The true measure of success lies in the intangible yet vital reconstruction of confidence, femininity, and the capacity for pleasure—reclaiming the narrative from one of loss to one of profound resilience and renewal. It is this imperative to heal the whole person—body, identity, and spirit—that defines the highest standard of care, ensuring every woman can look toward a future not just of survival, but of wholeness and pride.

AUTHORS' CONTRIBUTIONS

AP: Conceptualization, Formal Analysis, Methodology, Writing – original draft, Writing – review & editing. **JRMP:** Conceptualization, Investigation, Methodology, Resources, Supervision, Writing – review & editing. **BSM:** Data curation, Investigation, Visualization, Writing – original draft. **JRF:** Investigation, Resources, Validation, Writing – review & editing. **ECB:** Resources, Supervision, Validation, Writing – review & editing. **JMSJ:** Conceptualization, Resources, Supervision, Validation, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Costa MP, Ferreira MC, Soares JM Jr, Rossi AG, Baracat EC. Minimally invasive mastectomy: minimal incisions for better aesthetic quality of breast reconstruction. Eur J Gynaecol Oncol. 2012;33(2):155-8. PMID: 22611954
2. Nyirády LE, Czébely-Lénárt A, Hoferica J, Hegyi P, Fehérvári P, Nyirády P, et al. Evaluating oncoplastic breast-conserving surgery: oncological safety, risks, and satisfaction-a systematic review and meta-analysis. Sci Rep. 2025;16(1):444. https://doi.org/10.1038/s41598-025-30062-w
3. Veiga DF, Veiga-Filho J, Ribeiro LM, Archangelo I Jr, Balbino PF, Caetano LV, et al. Quality-of-life and self-esteem outcomes after oncoplastic breast-conserving surgery. Plast Reconstr Surg. 2010;125(3):811-7. https://doi.org/10.1097/PRS.0b013e3181ccdac5. Retraction in: J Cancer Res Ther. 2019;15(1):264. https://doi.org/10.4103/0973-1482.253971
4. Bachega FS, Turri JAO, Baracat MCP, Simões RS, Maciel GAR, Lobo RA, et al. New comprehension on polycystic ovary syndrome and sexual function: a systematic review and meta-analysis. J Sex Med. 2025;22(9):1612-28. https://doi.org/10.1093/jsxmed/qdaf163

https://doi.org/10.1590/1806-9282.20250706

Impact of uterine closure techniques following cesarean section on the development of cesarean scar defects: consigning cicatrix to history?

Ilker Sengul[1,2]*, Demet Sengul[3]

Dear Editor,

Cesarean section (C/S) scar defect (CSD), per se, has always occupied minds for a considerable time, which requires a graceful approach[1-7]. I am writing to you as an appraisal of the article entitled "Is cesarean scar defect becoming history? The effect of uterotomy closure" by Gumusburun and Uskent[8], published in volume 71, Rev Assoc Med Bras. This article presents a significant investigation into the impact of uterine closure techniques following C/S on the development of CSD, also known as isthmocele pathology. The primary strength of this study lies in its novel investigation of the "baseball" type suture technique in the context of isthmocele formation. Gumusburun and Uskent specifically evaluated the effects of this technique, comparing traditional single-layer locked and unlocked continuous suturing methods. The study's prospective design and the direct comparison of three distinct surgical techniques during primary C/S also contribute to its methodological rigor. The findings of the study are compelling, suggesting that the baseball-type closure significantly attenuated its incidence and led to a greater residual myometrial thickness (RMT) at the cicatrix site compared to the single-locked and single-unlocked techniques. Notably, isthmocele development was considerably lower in the baseball group compared to the single-locked and -unlocked ones. Furthermore, the highest mean RMT was observed in the baseball suture group, whose outcomes underscore the potential importance of the uterotomy closure technique in influencing cicatrix healing and preventing its formation. The authors rightly conclude that baseball-type suturing appears to be an advantageous method for preserving RMT and preventing isthmocele[8]. However, it is important to kindly acknowledge certain limitations of this study. The authors themselves mention that the statistical power was limited to 80%, and the postoperative sonographic evaluations were conducted only at the third month postpartum. While these findings are valuable, further studies with a higher power (e.g., 90%) and longer-term follow-up are needed to assess the long-term complications associated with that and to strengthen these initial observations. Moreover, the inclusion criteria for the study were quite restrictive, including only term pregnancies (>37 weeks) with no prior C/S, no active labor exceeding 4 cm cervical dilatation, and specific body mass index limits (<35 kg/m²). The exclusion of women with factors such as placental abnormalities, previous uterine surgery, multiple pregnancies, and certain comorbidities limits the generalizability of the findings to the broader population undergoing C/S. In fine, while this work meticulously details the three suture techniques used, it is noted that only one type of suture material was used across all groups. Therefore, future research could explore the potential impact of different suture materials on isthmocele formation. As a matter of fact, this study provides valuable initial evidence suggesting the superiority of the baseball-type uterine closure technique in reducing isthmocele pathology and improving RMT. This research recommends a promising avenue for refining surgical techniques during C/S to potentially mitigate the increasing prevalence of CSD and its associated complications. However, the identified limitations necessitate further investigation through larger, multicenter studies with long-term follow-up to validate these findings and to assess their applicability across a broader range of cases undergoing Cesarean delivery. This study serves as an important foundation for future research aimed at optimizing uterotomy closure techniques and potentially providing cesarean cicatrix defects,

[1]Giresun University, Faculty of Medicine, Division of Endocrine Surgery – Giresun, Turkey.
[2]Giresun University, Faculty of Medicine, Department of General Surgery – Giresun, Turkey.
[3]Giresun University, Faculty of Medicine, Department of Pathology – Giresun, Turkey.
*Corresponding author: ilker.sengul.52@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on April 11, 2025. Accepted on June 09, 2025.
Scientific Editor: José Maria Soares Júnior

a concern of the past. This issue merits further investigation. We thank Gumusburun and Uskent[8] for their valued study of uterotomy closure techniques in order to completely prevent CSD and consign it to history.

AUTHORS' CONTRIBUTIONS

IS: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **DS:** Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Sengul D, Sengul I, Soares Junior JM. Caesarean section scar endometriosis: quo vadis? Rev Assoc Med Bras (1992). 2022;68(1):1-2. https://doi.org/10.1590/1806-9282.20211074
2. Sengul I, Sengul D. Clinical presentation of incisional (cesarean section scar) endometriosis vary with the size of the lesion. J Clin Pract Res. 2011;33(2):171-2.
3. Sengul D, Sengul I. Scar endometriosis: a review. Remedy Open Access 2016;1(1):1038.
4. Gianfaldoni A, Roa C, Santos Simões R, Baracat MCP, Maggio Fonseca A, Bagnoli VR, et al. Association of intrauterine synechiae with pituitary gonadotrophin pulse patterns: a pilot study. PLoS One. 2023;18(12):e0289075. https://doi.org/10.1371/journal.pone.0289075
5. Sengul I, Sengul D, Kahyaoglu S, Kahyaoglu I. Incisional endometriosis: a report of 3 cases. Can J Surg. 2009;52(5):444-5. PMID: 19865584
6. Sengul I, Sengul D, Kahyaoğlu S, Kahyaoğlu I. Insizyonel endometriozis: uc olguluk bir sunu. Ulusal Cerrahi Kongresi 2008, Antalya, May 28-31, 2008. Abstract book GEN-15. p. 123.
7. Sengul I, Sengul D, Soares Junior JM. Interpretations on a rare localization of endometriosis: labium minus. Rev Assoc Med Bras (1992). 2021;67(1):1-2. https://doi.org/10.1590/1806-9282.67.01.001
8. Gumusburun N, Uskent U. Is cesarean scar defect becoming history? The effect of uterotomy closure. Rev Assoc Med Bras (1992). 2025;71(3):e20241507. https://doi.org/10.1590/1806-9282.20241507

https://doi.org/10.1590/1806-9282.20251818

Methodological concerns regarding "Use of high-dose steroid therapy: addition of anakinra in the treatment of severe COVID-19"

Fatih Yıldırım[1]*

Dear Editor,

We read with great interest the article titled "Use of high-dose steroid therapy: addition of anakinra in the treatment of severe COVID-19," recently published in your journal[1]. The study clearly reflects considerable clinical effort and dedication to understanding treatment approaches for severe COVID-19. However, several important methodological and statistical limitations may affect the validity and interpretation of its conclusions.

First, the hyperinflammatory state associated with COVID-19 differs from classical macrophage activation syndrome (MAS) in both clinical features and pathophysiology, as previously emphasized[2]. Labeling this hyperinflammatory state as MAS may lead to conceptual and clinical confusion, both in terms of nomenclature and diagnostic interpretation. The treatment of COVID-19–related hyperinflammatory states remains controversial, and no clear consensus has been reached in the literature regarding the efficacy of anti-cytokine therapies such as tocilizumab and anakinra. While early reports suggested that anakinra might improve hyperinflammation-related outcomes, subsequent systematic reviews could not confirm a mortality benefit[3,4].

When considering the methodological aspects of the study, several issues warrant attention. At baseline, the anakinra+steroid group exhibited significantly higher ferritin levels and more severe computed tomography (CT) findings, introducing potential selection bias, which can substantially distort treatment effect estimates in retrospective analyses. Although this important limitation was acknowledged within the article, matching or adjustment for baseline severity (e.g., CT severity score or oxygenation indices) could have strengthened the methodological robustness and comparability between treatment arms[5]. Moreover, variables such as malignancy or underlying rheumatic disease, which may directly affect mortality, were not adjusted for in regression models. These patient subgroups should either have been excluded at baseline or incorporated into the regression models. Some patients had likely received varying doses of corticosteroids and additional antimicrobial agents before initiation of high-dose steroid (250 mg) and anakinra therapy; these pre-treatments may have influenced treatment-related complications and mortality, and a brief discussion of this issue would have provided greater clarity.

Another important issue concerns pre-treatment oxygenation parameters (e.g., SpO_2, PaO_2/FiO_2, or need for high-flow oxygen). Although the limitations section of the article notes that the anakinra+steroid group had poorer oxygenation status before treatment, these parameters were neither presented in the tables nor included in the multivariable analysis. These are crucial covariates in assessing 28-day mortality, and their inclusion in the multivariable analysis might have yielded different results. Likewise, while the duration of steroid therapy was found to be significant in univariate analysis, it was not included in the multivariate model, which could substantially confound the main outcome. Evaluating both the duration of steroid therapy and the total steroid exposure in the multivariable analysis could have provided more robust and reliable results. Lastly, mortality analysis based solely on proportions without survival methods (e.g., Kaplan-Meier or Cox models) limits the interpretability of time-dependent outcomes[6].

In summary, the findings of this valuable study should be interpreted with caution in light of these limitations. Addressing the outlined methodological and statistical issues could significantly enhance the reliability and clinical applicability of the study's conclusions.

I have no conflict of interest with the authors of the original article that is the subject of this letter.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

[1]Çanakkale Mehmet Akif Ersoy State Hospital, Department of Rheumatology – Çanakkale, Turkey.
*Corresponding author: f08fatih@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on October 09, 2025. Accepted on October 26, 2025.
Scientific Editor: José Maria Soares Júnior

REFERENCES

1. Guclu KG, Geyiktepe-Guclu C, Bayramlar OF, Tuncer G, Aydin M. Use of high-dose steroid therapy: addition of anakinra in the treatment of severe COVID-19. Rev Assoc Med Bras (1992). 2024;70(1):e20230671. https://doi.org/10.1590/1806-9282.20230671
2. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ, et al. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet. 2020;395(10229):1033-4. https://doi.org/10.1016/S0140-6736(20)30628-0
3. Yildirim F, Erdogan M, Mutlu MY, Icacan OC, Onar MI, Bes C. Efficacy of anticytokine treatments added to corticosteroids in patients with COVID-19-associated pneumonia and hyperinflammation: a single center experience. Eur Rev Med Pharmacol Sci. 2022;26(19):7297-304. https://doi.org/10.26355/eurrev_202210_29924
4. Shang W, Zhang Y, Wang G, Han D. Anakinra was not associated with lower mortality in hospitalised COVID-19 patients: a systematic review and meta-analysis of randomized controlled trials. Rev Med Virol. 2023;33(2):e2418. https://doi.org/10.1002/rmv.2418
5. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res. 2011;46(3):399-424. https://doi.org/10.1080/00273171.2011.568786
6. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). BMJ. 1998;317(7172):1572. https://doi.org/10.1136/bmj.317.7172.1572

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20250931

The relationship between in-hospital mortality and Naples prognostic score in patients undergoing tricuspid valve surgery

Samet Sevinc[1]*, Yusuf Demir[2], Seda Tükenmez Karakurt[1], Hüseyin Karakurt[1], Salih Guler[3]

SUMMARY

OBJECTIVE: Predicting in-hospital mortality rates following tricuspid valve surgery is imperative, especially in light of the recent advancements in treatment modalities. The Naples prognostic score serves as an indicator of systemic inflammation, malnutrition, and prognostic outcomes across a spectrum of medical conditions. In this investigation, we sought to assess the efficacy of the Naples prognostic score in predicting in-hospital mortality among patients undergoing tricuspid valve surgery.

METHODS: A total of 360 consecutive patients diagnosed with severe tricuspid valve disease, who underwent tricuspid valve surgery, were included in this retrospective study. The study population was categorized based on the Naples prognostic score classification into two groups: low (0–1–2) and high (3–4).

RESULTS: The study's primary outcome—in-hospital mortality—was significantly higher in patients with a high Naples prognostic score group than in those with a low Naples prognostic score group (35 [28.2%] vs. 8 [5.1%], p<0.001). In the receiver operating characteristic analysis, the Naples prognostic score optimal cut-off value of >2 predicted in-hospital mortality with 81.4% sensitivity and 62.5% specificity (area under the curve: 0.782 [95%CI 0.729–0.829, p<0.0001]).

CONCLUSION: The findings of this study indicate that the Naples prognostic score serves as an independent predictor of in-hospital mortality among patients undergoing tricuspid valve surgery.

KEYWORDS: Mortality. Inflammation. Malnutrition.

INTRODUCTION

Tricuspid valve (TV) disease, particularly tricuspid regurgitation (TR), has garnered substantially less attention in comparison to aortic and mitral valve diseases, despite the frequency of TR's occurrence. This condition is frequently designated as the "forgotten valve"[1,2]. Despite the recommendations outlined in the guidelines, the surgical repair of the TV continues to be underutilized[3]. It has been reported that the in-hospital mortality rate for surgically treated TV diseases is 8.7%[4]. Presently, the logistic EuroSCORE, EuroSCORE II, and the Society of Thoracic Surgeons (STS) cardiac surgery risk model are not specifically tailored to forecast outcomes for these infrequent procedures. Although there have been some initiatives to develop a risk assessment model explicitly for tricuspid valve surgery (TVS), progress remains limited[5].

The Naples prognostic score (NPS) constitutes a significant scoring system that incorporates serum albumin, total cholesterol level, the neutrophil-lymphocyte ratio (NLR), and the lymphocyte-monocyte ratio (LMR). It functions as a valuable prognostic indicator across various cancer types[6].

A comprehensive literature review did not reveal any studies pertaining to the association between NPS and in-hospital mortality following TVS. This study seeks to investigate the predictive value of NPS for in-hospital mortality among patients undergoing TVS.

METHODS

A total of 360 consecutive patients diagnosed with severe TV disease who underwent TVS at our institution between February 2019 and December 2023 were included in this retrospective study. The exclusion criteria comprised patients with systemic inflammatory disorders, chronic autoimmune diseases, active systemic infections, those necessitating immediate surgical intervention for any reason, and individuals with

[1]Mehmet Akif Ersoy Thoracic and Cardiovascular Surgery Training and Research Hospital, Department of Cardiology – Istanbul, Turkey.
[2]Bakircay University, Cigli Training and Research Hospital – İzmir, Turkey.
[3]Mehmet Akif Ersoy Thoracic and Cardiovascular Surgery Training and Research Hospital, Department of Cardiovascular Surgery – Istanbul, Turkey.
*Corresponding author: sametsevincdr@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on August 23, 2025. Accepted on January 10, 2026.
Scientific Editor: José Maria Soares Júnior

a documented history of types 1–3 pulmonary arterial hypertension. Following the application of these exclusion criteria, a total of 280 patients were ultimately included in the study.

Prior to the procedure, blood samples were obtained from a forearm vein following a 12-h fasting period. The analysis of blood albumin levels was performed using a Cobas 8000 c502 analyzer (Roche Holding AG, Basel, Switzerland). Furthermore, the length of the hospital stay (LoS) was documented. The EuroSCORE II was computed for each patient (https://www.euroscore.org).

All subjects underwent preprocedural transthoracic echocardiography. A qualified echocardiographer conducted all echocardiographic assessments using a GE Vingmed Ultrasound AS instrument from Horten, Norway, equipped with a 3.2 MHz adult probe.

The Naples prognostic score comprises the following variables: serum albumin and total cholesterol levels, NLR, and LMR[6]. The Naples Prognostic Score (NPS) is based on four blood markers that assess a patient's nutritional status and systemic inflammation. The criteria for evaluating NPS are as follows:

1. NLR greater than 2.96
2. LMR less than 4.44
3. Total cholesterol (TC) at 180 mg/dL or above
4. Serum albumin levels below 4 g/dL

Each parameter contributes one point to the NPS, resulting in a total score that can range from 0 to 4. In this study, patients were classified as having a low NPS if their score was 0, 1, or 2, and as having a high NPS if their score was 3, 4.

The study protocol was approved by the local ethics committee and was conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all patients.

Statistical analyses

The baseline characteristics of the study population were divided according to NPS at admission; low (0–1–2) and high (3–4). The normality distribution of continuous variables was analyzed using analytical and visual methods. To determine the predictive value of the NPS value for the development of in-hospital mortality, receiver operating characteristic (ROC) curve was used. Effects of individual exposure were reported using hazard ratio (HR) and 95%CI. Univariate and multivariate logistic regression analyses were performed to determine the independent determinants of in-hospital mortality. The level of statistical significance was set at $p<0.05$, and an area under the ROC curve.

RESULTS

The baseline demographic, clinical, and procedural characteristics of the patients are summarized in Table 1. The study sample consisted of 280 patients with severe TV disease who underwent TVS. The mean age of the sample was 59±12 years. One hundred seventy-two (61.4%) patients were female. The sample was divided into two groups: low NPS and high NPS. There was no difference between the patients with a low NPS group and those with a high NPS group in age, presence of diabetes mellitus (DM), and pacemaker or implantable cardioverter defibrillators (ICD). In the High NPS group, female gender and atrial fibrillation were less common, while smoking was more prevalent. When echocardiographic parameters were evaluated, tricuspid annular plane systolic excursion and TV diameter were similar between the groups, while pulmonary arterial systolic pressure was higher and left ventricular ejection fraction was lower in the high NPS group. When evaluating laboratory parameters, levels of hemoglobin, albumin, total cholesterol, and lymphocyte counts were higher in the low NPS group, while neutrophil and monocyte counts, as well as creatinine levels, were elevated in the high NPS group. EuroSCORE II was higher in the high NPS group compared to the low NPS group.

In-hospital outcomes

The study's primary outcome—in-hospital mortality—was significantly higher in patients with a high NPS group than in those with a low NPS group (35 [28.2%] vs. 8 [5.1%], $p<0.001$). In parallel, the length of hospital stay (LoS) was significantly higher in patients with a high NPS than in those with a low NPS (Table 1).

Independent predictors of in-hospital mortality

Univariate logistic regression analysis revealed significant correlations between in-hospital mortality and age, hemoglobin, the presence of a pacemaker or ICD, atrial fibrillation, DM, creatinine, EuroSCORE II, and NPS (Table 2). Further analysis of these variables using the multivariate logistic regression analysis indicated that EuroSCORE II (HR 1.049, 95%CI 1–1.1; p=0.049), and NPS (HR 1.606, 95%CI 1.136–2.27; p=0.007), were independent predictors for the development of in-hospital mortality (Table 2).

A simple sampling of 1,000 samples was performed. The bootstrap results showed a EuroScore CI -0.23 to 1.68, p-value of 0.165, and an NPS CI 1.004 to 1.111, p-value of 0.04. Bootstrapping confirmed that the NPS remained significant.

In the ROC analysis, NPS optimal cut-off value of >2 predicted in-hospital mortality with 81.4% sensitivity and 62.5% specificity (area under the curve [AUC]: 0.782 [95%CI 0.729–0.829, p=0.003]), and EuroScore (AUC: 0.737 [95%CI 0.681–0.788, p=0.045]) (Figure 1).

Table 1. Demographic, clinical, and procedural characteristics of the study group.

	Low NPS (n=156)		High NPS (n=124)		All (n=280)		p
Age (years)	58	±12	61	±11	59	±12	0.116
Gender, n (%) (female)	109	69.9	63	50.8	172	61.4	**0.001**
Active smokers, n (%)	10	6.5	20	16.3	30	10.8	**0.009**
Patients with DM, n (%)	57	36.8	44	35.8	101	36.3	0.863
Pacemaker or ICD, n (%)	8	5.1	7	5.6	15	5.4	0.849
Atrial fibrillation, n (%)	129	82.7	89	71.8	218	77.9	**0.029**
TAPSE (mm)	17.1	±3.6	17.4	±3.9	17.2	±3.7	0.739
PASP (mmHg)	50	±15	54	±15	52	±15	**0.023**
Tricuspid valve diameter, mm	34.02	±13.2	34.24	±14.3	34.12	±13.7	0.416
LVEF (%)	55	±8	52	±10	54	±9	**0.003**
Hemoglobin level (g/L)	12.91	±1.76	11.91	±2.25	12.47	±2.05	**<0.001**
Monocyte count (10^3/mL)	0.57	±0.17	0.7	±0.23	0.63	±0.21	**<0.001**
Neutrophil count (10^3/mL)	4.37	±1.3	5.57	±1.81	4.9	±1.66	**<0.001**
Lymphocyte count (10^3/mL)	2.21	±0.66	1.52	±0.64	1.9	±0.7	**<0.001**
Creatinine level (mg/dL)	1	(0.71–1.04)	1.18	(0.74–1.2)	1.07	(0.71–1.09)	**0.001**
Albumin level (mg/dL)	4.25	±0.43	3.64	±0.65	3.98	±0.62	**<0.001**
Total cholesterol (mg/dL)	177.6	±46.6	143.9	±29.6	162.6	±43.3	**<0.001**
NLR	1.97	(1.53–2.6)	3.65	(2.96–5.13)	2.62	(1.74–3.57)	**<0.001**
LMR	3.96	(2.97–5.03)	2.11	(1.5–2.88)	3	(2.1–4.38)	**<0.001**
EuroSCORE II	2.46	(1.68–3.93)	3.46	(2.06–5.46)	2.8	(1.88–4.87)	**0.001**
LoS, (days)	11	(8–16)	13	(9–22)	12	(8–18)	**0.005**
In-hospital mortality, n (%)	8	5.1	35	28.2	43	15.4	**<0.001**

NPS: Naples prognostic score; DM: diabetes mellitus; ICD: implantable cardioverter defibrillators; TAPSE: tricuspid annular plane systolic excursion; PASP: pulmonary arterial systolic pressure; LVEF: left ventricular ejection fraction; NLR: neutrophil-lymphocyte ratio; LMR: lymphocyte-monocyte ratio; LoS: length of hospital stay. Values in bold are statistically significant at p<0.005.

Table 2. Results of the univariate and multivariate analyses of the variables in terms of their prognostic value in predicting in-hospital mortality in patients undergoing tricuspid valve surgery.

	Univariate analysis			Multivariate analysis		
	Univariate HR, 95%CI		p-value	Multivariate HR, 95%CI		p-value
EuroSCORE II	1.063	(1.014–1.114)	**0.011**	1.049	(1–1.1)	**0.049**
NPS	1.067	(1.147–2.25)	**0.006**	1.606	(1.136–2.27)	**0.007**

HR: hazard ratio; CI: confidence interval; p: probability statistic; NPS: Naples prognostic score. Values in bold are statistically significant at p<0.005.

DISCUSSION

This study's findings indicate that the high NPS is an independent predictor of in-hospital mortality in patients undergoing TVS. To the best of the authors' knowledge, this is the first study to investigate the prognostic value of the NPS in predicting in-hospital mortality for patients undergoing TVS. Our

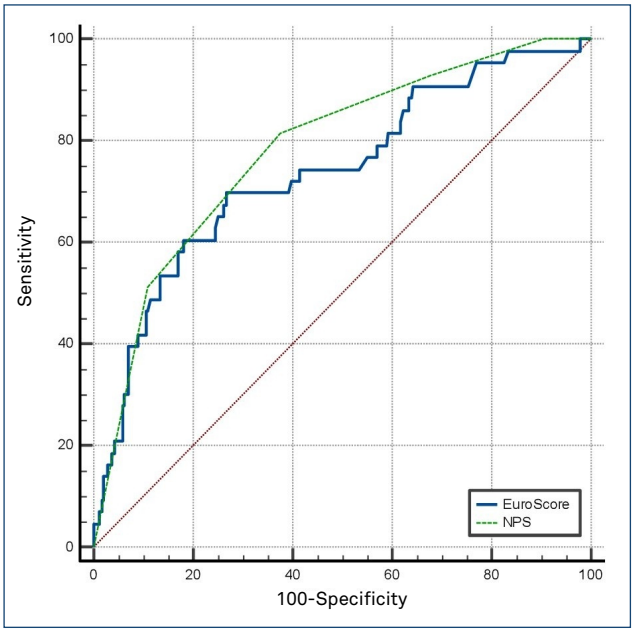


Figure 1. Receiver operating characteristic curve analysis of Naples prognostic score and EuroScore in patients undergoing tricuspid valve surgery.

results suggest that the NPS may be valuable in assessing an individual patient's mortality risk after TVS. The NPS is relatively simple to calculate, and its components often serve as hematological and biochemical indicators. By calculating the score early, we may be able to identify patients at higher risk for mortality, allowing us to take a more cautious approach in managing their care.

Notwithstanding its widespread occurrence and links to higher mortality and morbidity, surgery for TR is rarely conducted. Consequently, only 10% of patients diagnosed with TR receive surgical intervention[7]. Unlike aortic stenosis or mitral regurgitation, TR progresses slowly, with negative outcomes often emerging after many years, contributing to its misleadingly benign reputation. However, research consistently indicates a significant in-hospital mortality rate of 8–10%[5].

The NPS is an uncomplicated and readily accessible scoring system derived from serum albumin, total cholesterol, the NLR, and the LMR. In 2017, NPS was first utilized to assess the long-term survival of patients who underwent surgery for colorectal cancer[6]. Since then, researchers have examined the prognostic value of NPS in different cancers and chronic disease groups. A meta-analysis focusing on lung cancer patients revealed that a low NPS correlates with improved overall survival and disease-free survival rates[8]. Limited studies have explored the significance of NPS in predicting outcomes in cardiovascular diseases. Outstandingly, NPS has been shown to influence both in-hospital and midterm mortality rates, as well as rehospitalization among patients with functional class-3 and -4 heart failure resulting from decompensated heart failure[9]. In addition, Hakgor et al.'s study found that a high NPS was associated with a higher rate of in-hospital mortality following Transcatheter Aortic Valve Implantation[10].

There exists a limited number of studies in the literature that examine the relationship between inflammation and mortality in patients undergoing TV surgery. In a study conducted by Yoon et al., preoperative systemic inflammation, as assessed by the systemic immune-inflammation index, was found to be significantly associated with an elevated risk of postoperative composite complications and all-cause mortality within 1 month following TV surgery[11].

Although the link between hypoalbuminemia and poor surgical outcomes has been recognized for years, the underlying mechanisms remain unclear. Hypoalbuminemia, a marker of malnutrition, is associated with higher in-hospital and long-term mortality in patients before cardiac surgery[12]. This nutritional deficiency significantly heightens frailty in patients, negatively affecting their prognosis. Numerous scoring systems evaluating nutrition and physical status have been used in large-scale global studies, consistently linking functional and metabolic frailty to unfavorable outcomes[12]. Several reasons are proposed for this relationship. First, low serum albumin levels serve as an early warning sign for post-surgery outcomes, since preexisting inflammation can worsen the inflammatory response following surgery. Hypoalbuminemia often arises from chronic health conditions, with its severity reflecting the intensity of the inflammatory insult and associated mortality risk. Second, many studies link hypoalbuminemia to nutritional status, as albumin is a key indicator of in-hospital malnutrition and frailty. A low serum albumin level after surgery may indicate an unhealthy body mass index, which is connected to poor nutrition and inefficient metabolism, impairing the body's inflammatory and immune responses. Third, among elderly patients, hypoalbuminemia was common; serum albumin levels tend to decline with age, gradually decreasing from around 50 years old. Fourth, albumin's pharmacological roles as an antioxidant and transporter suggest that its deficiency might impair these functions, contributing to adverse outcomes[12]. In our study, it was observed that the in-hospital mortality rate was elevated among patients with TVS presenting lower levels of albumin, resulting in an increased NPS.

Neutrophils play a crucial role in managing inflammatory responses and releasing inflammatory mediators. When the inflammatory response is hyperactivated, it can negatively impact myocardial function due to maladaptive changes. This may result

in cardiac metabolic disorders, reduced contractility, ventricular dysfunction, and myocardial remodeling[13]. In contrast, lymphocytes serve to regulate the immune system's response. Notably, in patients with chronic heart failure, chronic inflammation, oxidative stress, and neurohormonal activation elevate plasma cortisol levels and increase catecholamine release. This process downregulates lymphocyte differentiation and proliferation, further leading to lymphocyte apoptosis[14]. Previous studies indicate that a low lymphocyte count can serve as an early marker of physiological stress[15]. In the study conducted by Shahim et al., it was shown that preoperative NLR predicted short- and long-term mortality after Aortic Valve Replacement surgery in patients diagnosed with severe AS[16]. It remains uncertain whether elevated NLR directly contributes to the development of TV diseases or if it is simply a marker indicating other systemic conditions that persist after TVS and raise mortality risks in the high NLR group. It is also unknown whether elevated NLR causes the increased mortality or merely reflects underlying etiological factors. If elevated NLR is connected to the development of TV diseases, the persistent higher mortality rates following TVS in the high NLR group imply that TV disease is part of a systemic process, not confined to TVS. Alternatively, elevated NLR might be a marker or component of a broader systemic abnormality with multiple effects. TV diseases could be one manifestation among others that develop over time post-TVS, contributing to higher mortality. Monocytes, which function as the precursors to macrophages, serve as indicators of the equilibrium between inflammation and immunity. The presence of preoperative monocytosis signifies insufficient activation of macrophages, leading to an imbalance favoring inflammation over immunity, thereby contributing to a less favorable prognosis for patients undergoing cardiac surgery. The LMR more effectively integrates the clinical significance of lymphocytes and monocytes. A lower LMR (less than 3.58) is associated with an increased risk of mortality among patients undergoing on-pump cardiac surgery, as evidenced by a cohort study with a 4-year follow-up period[17]. Our study observed that NPS increased in TVS patients with high NLR and low LMR, leading to a higher in-hospital mortality rate.

Serum cholesterol levels serve as indicators of nutritional status. Total cholesterol constitutes a complex lipid profile parameter that is associated with digestive functions, reflecting chronic inflammation and daily nutritional status[18]. Consequently, it may exert a direct influence on the primary endpoint of the study.

Moreover, the EuroSCORE and STS scores have not undergone validation for isolated TV surgery, as conventional scoring systems do not take into account the factors that contribute to heightened mortality in tricuspid operations. In our research, consistent with existing literature, a higher EuroSCORE II was observed in the high NPS group that underwent TVS and experienced elevated in-hospital mortality rates.

Elevated NPS scores were associated with a heightened risk of in-hospital mortality and LoS, irrespective of the presence of right ventricle dysfunction or underlying comorbidities. Consequently, the NPS may be employed to stratify the risk of patients undergoing TVS, potentially leading to improved surgical outcomes for TV procedures in the future.

Limitations

First, it was a retrospective, observational study conducted at a single center. Second, we only assessed baseline albumin, cholesterol, and hematological values; tracking changes through serial measurements could provide additional prognostic value. Additionally, we did not compare the NPS with traditional risk scoring systems, such as the TRI-SCORE, which is a new risk score designed to predict in-hospital mortality after isolated TV surgery[5]. Furthermore, relying solely on ROC analysis to identify the optimal NPS cut-off is a methodological limitation of the study.

CONCLUSION

This study represents the first comprehensive examination of the utility of the NPS as a preoperative risk assessment tool for patients undergoing TVS. This study's findings show that the NPS independently predicts in-hospital mortality for patients undergoing TVS.

ETHICAL APPROVAL

The study protocol was approved by the local ethics committee and was conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all patients.

AUTHORS' CONTRIBUTIONS

SS: Conceptualization, Data curation, Investigation, Writing – original draft, Writing – review & editing. **YD:** Conceptualization, Methodology, Writing – review & editing. **STK:** Investigation, Supervision. **HK:** Investigation. **SG:** Formal Analysis, Investigation.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Boudoulas KD, Barbetseas J, Pitsis AA, Boudoulas H. Tricuspid valve disease: the "forgotten valve". Cardiology. 2019;142(4):235-8. https://doi.org/10.1159/000497816
2. Braunwald NS, Ross J, Morrow AG. Conservative management of tricuspid regurgitation in patients undergoing mitral valve replacement. Circulation. 1967;35(4 Suppl):I63-9. https://doi.org/10.1161/01.cir.35.4s1.i-63
3. Shimoda TM, Ueyama HA, Miyamoto Y, Watanabe A, Gotanda H, Elmariah S, et al. Outcomes of isolated tricuspid replacement versus repair among older patients with tricuspid regurgitation in the United States. JTCVS Open. 2024;24:127-46. https://doi.org/10.1016/j.xjon.2024.10.018
4. Kawsara A, Alqahtani F, Nkomo VT, Eleid MF, Pislaru SV, Rihal CS, et al. Determinants of morbidity and mortality associated with isolated tricuspid valve surgery. J Am Heart Assoc. 2021;10(2):e018417. https://doi.org/10.1161/JAHA.120.018417
5. Dreyfus J, Bohbot Y, Coisne A, Lavie-Badie Y, Flagiello M, Bazire B, et al. Predictive value of the TRI-SCORE for in-hospital mortality after redo isolated tricuspid valve surgery. Heart. 2023;109(12):951-8. https://doi.org/10.1136/heartjnl-2022-322167
6. Galizia G, Lieto E, Auricchio A, Cardella F, Mabilia A, Podzemny V, et al. Naples prognostic score, based on nutritional and inflammatory status, is an independent predictor of long-term outcome in patients undergoing surgery for colorectal cancer. Dis Colon Rectum. 2017;60(12):1273-84. https://doi.org/10.1097/DCR.0000000000000961
7. Messika-Zeitoun D, Candolfi P, Dreyfus J, Burwash IG, Iung B, Philippon JF, et al. Management and outcome of patients admitted with tricuspid regurgitation in France. Can J Cardiol. 2021;37(7):1078-85. https://doi.org/10.1016/j.cjca.2020.12.012
8. Wang YS, Niu L, Shi WX, Li XY, Shen L. Naples prognostic score as a predictor of outcomes in lung cancer: a systematic review and meta-analysis. Eur Rev Med Pharmacol Sci. 2023;27(17):8144-53. https://doi.org/10.26355/eurrev_202309_33574
9. Erdogan A, Genc O, Inan D, Yildiz U, Balaban I, Guler Y, et al. Impact of Naples prognostic score on midterm all-cause mortality in patients with decompensated heart failure. Biomark Med. 2023;17(4):219-30. https://doi.org/10.2217/bmm-2022-0689
10. Hakgor A, Dursun A, Kahraman BC, Yazar A, Savur U, Akhundova A, et al. The impact of the Naples prognostic score on the short- and long-term prognosis of patients undergoing transcatheter aortic valve implantation. J Cardiovasc Med (Hagerstown). 2024;25(7):519-28. https://doi.org/10.2459/JCM.0000000000001637
11. Yoon J, Jung J, Ahn Y, Oh J. Systemic immune-inflammation index predicted short-term outcomes in patients undergoing isolated tricuspid valve surgery. J Clin Med. 2021;10(18):4147. https://doi.org/10.3390/jcm10184147
12. Xu R, Hao M, Zhou W, Liu M, Wei Y, Xu J, et al. Preoperative hypoalbuminemia in patients undergoing cardiac surgery: a meta-analysis. Surg Today. 2023;53(8):861-72. https://doi.org/10.1007/s00595-022-02566-9
13. Sionis A, Suades R, Sans-Roselló J, Sánchez-Martínez M, Crespo J, Padró T, et al. Circulating microparticles are associated with clinical severity of persistent ST-segment elevation myocardial infarction complicated with cardiogenic shock. Int J Cardiol. 2018;258:249-56. https://doi.org/10.1016/j.ijcard.2017.10.044
14. Barron HV, Cannon CP, Murphy SA, Braunwald E, Gibson CM. Association between white blood cell count, epicardial blood flow, myocardial perfusion, and clinical outcomes in the setting of acute myocardial infarction: a thrombolysis in myocardial infarction 10 substudy. Circulation. 2000;102(19):2329-34. https://doi.org/10.1161/01.cir.102.19.2329
15. Shumilah AM, Othman AM, Al-Madhagi AK. Accuracy of neutrophil to lymphocyte and monocyte to lymphocyte ratios as new inflammatory markers in acute coronary syndrome. BMC Cardiovasc Disord. 2021;21:422. https://doi.org/10.1186/s12872-021-02236-7
16. Shahim B, Redfors B, Lindman BR, Chen S, Dahlen T, Nazif T, et al. Neutrophil-to-lymphocyte ratios in patients undergoing aortic valve replacement: the PARTNER trials and registries. J Am Heart Assoc. 2022;11(11):e024091. https://doi.org/10.1161/JAHA.121.024091
17. Zhou Z, Liang M, Wu H, Huang S, Weng R, Hou J, et al. Preoperative lymphocyte-to-monocyte ratio as a prognostic predictor of long-term mortality in cardiac surgery patients: a propensity score matching analysis. Front Cardiovasc Med. 2021;8:639890. https://doi.org/10.3389/fcvm.2021.639890
18. Omran ML, Morley JE. Assessment of protein energy malnutrition in older persons, Part II: laboratory evaluation. Nutrition. 2000;16(2):131-40. https://doi.org/10.1016/s0899-9007(99)00251-8

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251229

Efficacy of a virtual multimodal exercise program on postural awareness and digital eye strain in university students: a pilot study

Ayça Aracı[1], Havva Nur Mutlu[1], Selenay Akgün[1], Ali Kurt[2]*

SUMMARY

OBJECTIVE: Excessive screen exposure is associated with musculoskeletal discomfort, postural alterations, visual fatigue, and increased anxiety among young adults. This study investigated the effects of a virtually delivered multimodal exercise program integrating sensorimotor, oculomotor, and postural components on postural awareness, musculoskeletal complaints, visual fatigue, and anxiety among screen-exposed university students and staff.

METHODS: Fifty-seven participants (mean age: 23.35±1.2 years) with ≥6 h of daily screen use completed an 8-week program consisting of twice-weekly live virtual sessions and daily video-guided exercises. Outcomes included the Postural Habits and Awareness Scale, Cornell Musculoskeletal Discomfort Questionnaire, Rapid Office Strain Assessment, ASTHENOPIA Eye Fatigue Scale, and State-Trait Anxiety Inventory. Pre–post differences were analyzed using paired-sample t-tests and Wilcoxon signed-rank tests.

RESULTS: Significant improvements were observed in Postural Habits and Awareness Scale Factor 1 and Factor 3 ($p<0.05$), accompanied by reduced musculoskeletal discomfort across all Cornell Musculoskeletal Discomfort Questionnaire sections ($p<0.05$). Visual fatigue decreased substantially ($z=-4.185$; $p<0.001$), whereas Rapid Office Strain Assessment scores showed no significant change ($p>0.05$). State anxiety increased slightly ($p<0.05$), while trait anxiety remained stable.

CONCLUSION: This preliminary single-group study suggests that virtual multimodal exercise programs may be effective, accessible preventive strategies for reducing musculoskeletal symptoms, enhancing posture awareness, and alleviating visual fatigue among individuals with high daily screen exposure. Controlled trials with larger and more diverse populations are recommended to establish causal effectiveness.

KEYWORDS: Exercise therapy. Posture. Asthenopia. Anxiety. Students.

INTRODUCTION

The rapid digitalization of academic and professional environments has substantially increased daily exposure to screen-based technologies, particularly among university students and staff who rely on digital platforms for coursework, communication, and administrative tasks. Recent evidence suggests that prolonged screen exposure is strongly linked to musculoskeletal discomfort, alterations in postural alignment, and decreased overall well-being[1-3]. Repetitive upper-extremity activity, static seated postures, and suboptimal workstation habits contribute to the growing prevalence of musculoskeletal symptoms in digitally engaged populations.

Beyond physical consequences, excessive screen exposure is closely linked to oculomotor fatigue, visual discomfort, attentional instability, and digital eye strain (DES), a syndrome whose prevalence has increased sharply over the past decade[4-7]. Digital media use has also been shown to impair sleep patterns and elevate psychological stress, underscoring the multidimensional health impact of intensive screen engagement[5,8]. These emerging findings highlight the complex interplay between visual, cognitive, and musculoskeletal demands in digital environments.

Increasing evidence suggests that extended screen use disrupts the integration of multisensory systems, particularly visual, vestibular, and proprioceptive inputs that are essential for maintaining postural control and sensorimotor regulation[2,9,10]. Altered cervical proprioception and diminished head-eye coordination, frequently observed in screen-exposed individuals, may impair reflexive pathways such as the cervico-ocular and cervicocollic reflexes, contributing to balance instability and sensorimotor inefficiency. Despite these risks, multimodal interventions addressing the multisensory consequences of prolonged screen exposure remain limited.

While prior studies have examined isolated approaches such as general stretching, visual exercises, or physical activity programs, few have investigated virtual or hybrid interventions that simultaneously target sensorimotor, oculomotor, and postural domains[3,11]. Given the rapid evolution of remote learning, hybrid work, and digital communication platforms, there is a growing need for accessible, evidence-based strategies that address

[1]Alanya Alaaddin Keykubat University, Faculty of Health Science, Department of Physiotherapy and Rehabilitation – Alanya, Türkiye.
[2]Alanya Alaaddin Keykubat University, Faculty of Medicine, Department of Ophthalmology – Alanya, Türkiye.
*Corresponding author: dralikurt@gmail.com

Conflicts of interest: the authors declare there is no conflicts of interest. Funding: this study was funded by TUBITAK (The Scientific and Technological Research Council of Turkey), Project No: [1919B012324064].
Received on August 27, 2025. Accepted on December 29, 2025.
Scientific Editor: José Maria Soares Júnior

the interconnected nature of postural awareness, proprioceptive acuity, and oculomotor function[7,11,12].

However, the literature remains limited regarding multimodal virtual interventions that integrate sensorimotor, oculomotor, and postural training within real-time supervised digital platforms. Most existing studies have focused on single-component exercise approaches rather than comprehensive and integrated protocols, which limits their applicability to the multidimensional demands of prolonged screen-based activity[7,11,12].

Therefore, the present study aimed to evaluate the effects of a virtually delivered multimodal exercise intervention, including sensorimotor, oculomotor, and postural components, on postural awareness, musculoskeletal discomfort, visual fatigue, and anxiety among university students and staff with high daily screen exposure. By addressing a multifaceted health concern with an integrated approach, this study aims to contribute updated and clinically relevant evidence to support preventive strategies that mitigate the physical and psychological consequences of digital technology use.

METHODS

Study design and setting

This cross-sectional descriptive study included young adults (students and staff) from Alanya Alaaddin Keykubat University, selected through convenience sampling based on voluntary participation. The sample's homogeneity was intentional, reflecting the practical scope of a TÜBİTAK 2209 student research project. Accordingly, demographic diversity was limited in terms of age, occupation, and education level. This controlled homogeneity aimed to enhance internal validity by minimizing the influence of extraneous variables such as occupational differences, ergonomic diversity, and age-related variability in sensorimotor function. Nevertheless, we recognize that such intentional homogeneity restricts generalizability to wider screen-exposed adult populations.

Eligibility criteria

The inclusion criteria included being aged between 18 and 65 years, having a minimum of 6 h per day of cumulative screen exposure, including computer, smartphone, and tablet use for academic, professional, or recreational purposes.

Exclusion criteria were a history of spinal surgery or trauma, neurological deficits, vestibular disorders, any cardiopulmonary or musculoskeletal conditions affecting physical performance, and pregnancy.

Recruitment: Participants completed an 8-week program that included twice-weekly live sessions (20–25 min each) via WhatsApp Web and a 15-min pre-recorded video comprising the exercises listed in Table 1. The video was shared with

Table 1. Exercise protocols and delivery parameters.

Step	Description	Type
Group 1: Oculomotor sensorimotor exercises		
1	Head in neutral, fixate on colored object in right hand and count to 10, then switch to object in left hand and count to 10.	Fixation exercise
2	Head in neutral, follow colored object in right hand as it moves from right to left and back without moving head.	Smooth pursuit
3	With colored object held to the left, fixate with eyes while rotating the head right and left.	Adaptation X1
4	Fixate on colored object and rotate arm and head in opposite directions.	Adaptation X2
Group 2: Postural exercises		
1	Wrist and finger stretch with palms outward, arms overhead while keeping elbows extended. Hold for 5 s, repeat three times.	Stretching
2	Shoulder girdle stretch by pulling one arm medially behind the head using the opposite hand. Include lateral trunk flexion. Repeat bilaterally three times.	Stretching
3	With hands on lower back, press lumbar spine forward while extending neck. Then cross legs, rotate trunk and head toward upper leg. Repeat three times bilaterally.	Stretching
4	Move both arms backward: one at 135° flexion (palm forward) and the other at 45° flexion (back of hand forward). Hold for 5 s. Repeat bilaterally three times.	Stretching
5	Elevate shoulders as high as possible and then slowly depress while keeping elbows extended. Repeat three times.	Scapular Elevation/ Depression
6	With arms at 90° abduction and elbows extended, bring arms slowly together to 90° flexion and return. Repeat three times.	Scapular abduction/ adduction
7	Unilateral trapezius stretch by stabilizing upper trapezius with opposite hand and laterally flexing the head. Hold for 5 s, repeat three times bilaterally.	Trapezius stretch
8	Diaphragmatic breathing: Place left hand on chest and right hand on upper abdomen. Breathe in through the nose, inflating abdomen without chest movement. Repeat five times.	Breathing exercise

Participants completed an 8-week program including twice-weekly live sessions (20–25 min each) via WhatsApp Web and a 15-min pre-recorded video composed of the exercises listed in Table 1. The video was shared with participants daily for self-practice. All sessions were instructed and demonstrated by a certified physiotherapist (MSc) specializing in musculoskeletal rehabilitation. Participants were encouraged to repeat the video session at least three times per day and to record their adherence through weekly feedback forms. Based on adherence logs, participants completed a mean of 2.4±0.8 daily video repetitions and attended 86% of all scheduled live sessions, yielding an overall adherence rate of 82% across the 8-week program.

participants via WhatsApp for daily self-practice. Exercises were instructed and demonstrated by a certified physiotherapist specializing in musculoskeletal rehabilitation. Participants were encouraged to repeat the video session at least three times a day and to record their adherence through weekly feedback forms. The overall attendance rate for live sessions was approximately 88%. Based on weekly adherence logs, participants completed a mean of 2.4±0.8 daily video repetitions and attended 86% of all scheduled live sessions, yielding an overall adherence rate of 82% across the 8-week program (Table 1).

The data collection stage was conducted using a structured evaluation form administered within an internet environment. The evaluation form was sent via a Google Forms link over various communication routes (Messaging, WhatsApp, etc.) and social media platforms (Facebook, Instagram, etc.).

Assessment and data collection

Demographic Information and Eye Fatigue Assessment: Data were collected through a structured, three-part online questionnaire distributed via Google Forms. The first section captured demographic data (e.g., age, gender) and visual environment factors such as daily screen time, device type, and use of glasses or screen filters.

The second section assessed eye-related symptoms (e.g., burning, dryness, blurred vision, headache) experienced after 1 h of screen exposure, rated on a five-point frequency scale from "never" to "often." The final section evaluated symptom changes during the online education period, with responses ranging from "decreased a lot" to "increased a lot"[13].

Postural Habits and Awareness Scale

The PHAS, developed by Bayar et al. and validated in Turkish, evaluates individuals' postural habits and awareness in daily activities. The scale consists of 19 items rated on a five-point Likert scale ("1=strongly disagree" to "5=strongly agree"), with seven reverse-scored items. It comprises two subscales: postural habits (max score=35) and postural awareness (max score=60), totaling a maximum score of 95. Higher scores reflect better posture and greater awareness in activities such as standing, sitting, lying, shopping, and lifting[14].

Cornell Musculoskeletal Disorders Questionnaire (CMDQ): The CMDQ, validated in Turkish by Erdinc et al., was used to assess musculoskeletal symptoms over the past week[15]. It evaluates the frequency, intensity, and work-related impact of pain or discomfort in specific body regions. Participants rated how often they experienced symptoms, their severity, and whether these affected their work. Individual risk scores were calculated based on their responses.

Rapid Office Strain Assessment (ROSA): The ROSA tool was used to evaluate participants' ergonomic risk levels related to office workstation design. This observational instrument assesses postural risks associated with chairs, desks, screens, keyboards, mice, and telephones, based on established methods such as Rapid Upper Limb Assesment (RULA) and Rapid Entire Body Assesment (REBA)[16].

In this study, ROSA was included as a secondary outcome measure to monitor potential variations in ergonomic risk during the intervention. Because the exercise program did not involve any physical modification of workstation components, a substantial change in ROSA scores was not expected.

Each participant's workstation was assessed on-site by a trained observer using the standardized ROSA diagrams and matrix. Risk scores were calculated across four domains:

- Chair: Adjustability and lumbar support features
- Screen and phone: Alignment, lighting, and usage duration
- Keyboard and mouse: Posture and frequency of use
- Overall risk: Combined total of all components

Higher ROSA scores indicate elevated ergonomic risk. Following the reviewer's recommendation, the manuscript now explicitly presents the risk thresholds defined by Sonne et al.[16]:

- ≥7: High ergonomic risk
- 5–7: Moderate risk
- 0–4: Low risk

Spielberger State and Trait Anxiety Inventory

The STAI was used to assess participants' anxiety levels, consisting of two subscales: STAI-I (state anxiety) and STAI-II (trait anxiety), each with 20 items. Both forms include direct and reverse-coded items, and scoring is based on a standardized formula. Higher total scores reflect greater anxiety. The Turkish adaptation of the STAI was validated between 1974 and 1977[17] and it remains a widely accepted tool for evaluating emotional state variability in psychological research.

Statistical analysis: Statistical analysis was performed using the Statistical Package for the Social Sciences (SPSS), version 25.0 (IBM Corp., Armonk, NY, USA). The normality of the data distribution was evaluated using the Shapiro-Wilk test. Paired samples t-test was used for normally distributed variables, and a Wilcoxon signed-rank test was used for non-normally distributed variables. Demographic findings, clinical findings, and delta (Δ) values were reported using descriptive statistics, including mean±standard deviation for normal distribution data and median (interquartile range) for the data that has not been normally distributed. The delta (Δ) values are calculated from the differences between the pre-test and post-test. A significance level of $p<0.05$ was accepted in all analyses.

Ethics statement: Ethical approval was obtained from the Alanya Alaaddin Keykubat University Non-Interventional Clinical Research Ethics Committee (09/07/2024, Decision No: 2024/05), and the study was conducted in accordance with the Declaration of Helsinki. The study was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK 2209-A, Project No: 1919B012324064).

Participants were informed about the study through a WhatsApp group meeting, after which they received consent forms via an online platform. Both verbal and written consents were obtained prior to the commencement of data collection.

RESULTS

A total of 57 volunteer participants were included in the study. The study comprised 70.2% female participants (n=40) and 29.8% male participants (n=17). The mean age of the participants was 23.35±1.2 years, with an age range of 22–27 years. The mean height was 168.63±8.22 cm, body weight was 64.84±13.14 kg, and body mass index (BMI) was 22.65±3.27 (Table 2).

All participants were university students, and their education level was undergraduate (100%). Regarding their computer use over the last 6 months, 26.3% stated that they had never used a computer, 47.4% reported using a computer for less than 1 h a day, and 26.3% stated that they used a computer for 2 h or more a day. The average duration of computer use was reported as 3.88±2.78 years. The average number of working days per week was 3.95±2.86 days. Sixty-four point nine percent of the participants stated that they took breaks during computer use (Table 2).

In terms of exercise habits, 26.3% exercised regularly several times a week, 59.6% exercised occasionally, and 14.0% did not exercise at all. In terms of smoking habits, 40.4% had never smoked, 17.5% had smoked before but quit, and 42.1% were current smokers (Table 2).

All participants (100%) had at least one chronic disease. The most frequently reported chronic diseases were headache (45.6%), eye diseases (26.3%), cervical hernia (8.8%), hypertension (7.0%), eczema (7.0%), diabetes (1.8%), thyroid diseases (1.8%), and lumbar hernia (1.8%) (Table 2).

No significant change was observed in any of the dimensions of the ROSA scale (p>0.05) (Table 3).

Table 2. Demographic information.

		n	%
Gender	Woman	40	70.2
	Man	17	29.8
Age	Mean±SD	23.35±1.2	
	Med (IQR)	23 (22.5–24)	
	Min–max	22–27	
Weight	Mean±SD	64.84±13.14	
	Med (IQR)	62 (55.5–72.5)	
	Min–max	42–97	
Height	Mean±SD	168.63±8.22	
	Med (IQR)	166 (163–174.5)	
	Min–max	149–188	
BMI	Mean±SD	22.65±3.27	
	Med (IQR)	22.66 (19.55–25.54)	
	Min–max	16.8–29.74	
Level of education	University	57	100.0
Have you been using a computer at home in the last 6 months?	No	15	26.3
	Yes, less than 1 h a day	27	47.4
	2 h or more than 2 h	15	26.3
Computer use duration of more than 1 h per day (in years)	Mean±SD	3.88±2.78	
	Med (IQR)	4 (3–5)	
	Min–max	0–20	

Continue...

Table 2. Continuation.

		n	%
How many days a week have you been working in the past 6 months?	Mean±SD	4.51±2.34	
	Med (IQR)	4 (4–5)	
	Min–max	2–20	
Do you take breaks while using the computer?	No	20	35.1
	Yes	37	64.9
Do you exercise?	I do it regularly several times a week	15	26.3
	Occasionally	34	59.6
	No	8	14.0
Do you smoke	I have never smoked	23	40.4
	I used to smoke, but I quit	10	17.5
	I currently smoke	24	42.1
Do you have any chronic illnesses?	Yes	57	100.0
Which chronic illness do you have?	Headache	26	45.6
	Hypertension	4	7.0
	Eczema	4	7.0
	Diabetes	1	1.8
	Eye diseases	15	26.3
	Thyroid disorders	1	1.8
	Cervical disc herniation	5	8.8
	Lumbar disc herniation	1	1.8

n: frequency; %: percent; SD: standard deviation; BMI: body mass index; Med (IQR): median (25th–75th percentiles); min–max: minimum–maximum values.

Table 3. Within-group differences in pre- and post-test scores.

	Pre-test	Post-test	Intra group p	Delta Δ
Rosa A	5 (4–5)	5 (4–5)	0.414 (z=-0.816)	0 (0–0)
Rosa monitor B	1 (1–1)	1 (1–1)	1 (z=0)	0 (0–0)
Rosa C	4 (4–4)	4 (4–4)	0.48 (z=-0.707)	0 (0–0)
Rosa toplam	10 (9–10)	10 (9–10)	0.285 (z=-1.069)	0 (0–0)
PHAS FACTOR 1	24.74±3.9	25.77±3.83	**0.012* (t=-2.585)**	-1 (-4–1)
PHAS FACTOR 2	15 (14–16)	16 (12–16)	0.682 (z=-0.41)	0 (-2–2)
PHAS FACTOR 3	12 (11–14)	14 (12–15)	**0.0001* (z=-4.148)**	-2 (-3–0)
PHAS FACTOR 4	8.28±1.92	7.89±2.01	0.199 (t=1.3)	0 (-1–2)
CMDQ Part 1	43 (37.5–52)	34 (25–43.5)	**0.0001* (z=-6.576)**	9 (7.5–12)
CMDQ Part 2	46 (41–52.5)	40 (32.5–45.5)	**0.0001* (z=-6.256)**	6 (2.5–8)
CMDQ Part 3	33.49±8.22	27.51±8.38	**0.0001* (t=14.865)**	6 (4–8)
STAI-I State Anxiety	47 (45–49)	48 (46–50)	**0.005* (z=-2.804)**	-1 (-2–0)
STAI-II Trait Anxiety	47 (46–49)	47 (45–48.5)	0.301 (z=-1.034)	1 (-1–2)
ASTHENOPIA	1 (0–2)	0 (0–1.5)	**0.001* (z=-4.185)**	0 (0–1)

*p<0.05 indicates a statistically significant change and is represented in bold; SD: standard deviation; Med (IQR): median (25th–75th percentiles); z: Wilcoxon Signed-Rank Test. Normally distributed variables were analyzed with paired-samples t-test and presented as mean±SD. Non-normally distributed variables were analyzed with Wilcoxon signed-rank test and presented as median (IQR). Delta (Δ) values were reported accordingly. PHAS: Postural Habits and Awareness Scale; CMDQ: Cornell Musculoskeletal Discomfort Questionnaire; STAI-I: State-Trait Anxiety Inventory-I; STAI-II: State-Trait Anxiety Inventory-II.

A significant increase was observed in the sub-dimensions of Posture Habit and Awareness (Factor 1) and Positional Awareness (Factor 3) ($p<0.05$), indicating that participants experienced positive improvements in both posture awareness and postural position awareness in their daily lives. No significant change was found in the Awareness of Factors Disturbing Posture (Factor 2) and Ergonomic Awareness (Factor 4) sub-dimensions ($p>0.05$) (Table 3).

A significant decrease was observed in all sections of the CMDQ ($p<0.05$), indicating a decrease in musculoskeletal complaints among the participants (Table 3).

There was a significant increase in the Spielberger STAI ($p<0.05$), whereas no significant change was observed in the trait anxiety level ($p>0.05$) (Table 3).

Eye fatigue showed a statistically significant reduction following the intervention (pre: 1.79±2.41; post: 1.19±2.13), confirmed by the Wilcoxon test ($z=-4.185$, $p<0.001$). Median scores decreased from 1.00 (0.00–2.00) to 0.00 (0.00–1.50) (Table 3).

These values correspond exactly to those reported in Table 3.

DISCUSSION

This study demonstrated that a virtually delivered multimodal exercise program, combining sensorimotor, oculomotor, and postural components, produced significant improvements in postural awareness, musculoskeletal complaints, and visual fatigue among university students and staff who were exposed to prolonged daily screen use. The marked reductions in Posture Habit and Awareness (Factor 1), Positional Awareness (Factor 3), and CMDQ scores are consistent with recent findings that short, structured multimodal exercise programs can effectively enhance proprioceptive efficiency, correct postural habits, and reduce musculoskeletal symptom burden in digitally engaged populations[2,3]. Following the widespread adoption of virtual learning and hybrid work models, simple online feedback-based exercise routines have also been shown to improve alignment of the head, shoulders, and trunk[1-3,11] These findings collectively highlight the potential of virtual sensorimotor-based interventions as accessible strategies for addressing the biomechanical consequences of prolonged screen exposure.

The significant reduction in Asthenopia scores ($z=-4.185$, $p<0.001$) provides robust evidence that the oculomotor and sensorimotor components of the virtual program effectively alleviated visual fatigue. This finding aligns with recent studies demonstrating that targeted oculomotor training reduces DES by improving accommodative efficiency, blink regularity, and ciliary muscle control[4-7]. Similar improvements in visuocognitive and oculomotor task performance following virtual reality–based training have also been demonstrated in both peri- and post-menopausal adults, highlighting the neuroplastic potential of visually guided digital interventions[18]. Given the growing prevalence of screen-related visual symptoms among young adults, the improvement observed in our study reinforces the value of accessible, virtual eye–neck coordination exercises in mitigating visual discomfort[1,2].

The absence of significant changes in ROSA scores should be interpreted within the context of both the study population and the inherent characteristics of the ROSA assessment tool. Most participants were young adults whose daily routines did not primarily involve sustained desk-based work. Consequently, their cumulative ergonomic exposure was limited, reducing the likelihood of measurable changes over the short intervention period. ROSA was originally designed to evaluate static workstation configurations and sustained postural demands typical of full-time office workers[6,19]. In this population, baseline ROSA values (mean≈9.5) reflected suboptimal workstation setups rather than accumulated ergonomic strain.

Moreover, ROSA primarily quantifies equipment-related and layout-based ergonomic parameters, with limited sensitivity to behavioral or perceptual adaptations, which are the specific targets of virtual sensorimotor–postural training. Thus, improvements in postural awareness identified through PHAS may not translate into short-term changes on a tool focused on static workstation features. Future studies may benefit from combining ROSA with subjective ergonomic perception scales or digital posture-tracking systems to capture subtle changes in ergonomic behavior among individuals with low-intensity workstation exposure.

The unexpected increase in state anxiety (STAI-I) warrants careful consideration. Several mechanisms, supported by emerging evidence, may explain this response. Virtual and hybrid learning environments are known to increase cognitive load, attentional demand, and self-focused monitoring, contributing to short-term elevations in anxiety[8,12]. Virtual simulator-based studies have further demonstrated that visual distraction and increased attentional load significantly influence motor reaction time and postural responses, indicating that visually demanding digital tasks can induce heightened neurocognitive strain[20]. Live group-based virtual sessions may amplify performance-related self-awareness and social evaluative concerns, particularly when participants feel observed by peers[21].

Additionally, early engagement in body-focused, awareness-based, and sensorimotor practices can temporarily heighten interoceptive awareness, a mechanism shown to intensify anxiety during the initial stages of posture- and awareness-based interventions[7,9,10]. This response may be further amplified in visually and cognitively engaging digital environments. A recent

systematic review of virtual reality-based interventions during midlife reported consistent enhancements in visuomotor coordination and cognitive engagement, highlighting the increased attentional and interoceptive demands imposed by immersive digital exercise platforms[22]. For young adults, who may be less familiar with interoceptive training, increased internal attention can magnify uncertainty regarding task performance and amplify transient anxiety responses.

Digital fatigue may also contribute. Individuals with high baseline screen exposure may experience an added mental load during virtual sessions, consistent with meta-analytic evidence linking videoconferencing to "Zoom fatigue" and elevated physiological[10] arousal[8,12]. Importantly, anxiety scores remained below clinically significant thresholds and did not affect adherence, indicating a transient and manageable response rather than an adverse psychological outcome.

Future interventions may reduce anxiety reactivity by incorporating structured onboarding, adaptive difficulty progression, real-time guidance, and more opportunities for clarifying expectations and performance criteria. Tracking engagement levels and integrating qualitative feedback may further enhance psychological safety.

Collectively, the results demonstrate that virtual multimodal exercise programs represent a promising, accessible, and multidimensional approach to addressing the physical and visual demands of digital environments. By simultaneously targeting sensorimotor, oculomotor, and postural pathways, these interventions may serve as effective tools for mitigating screen-related musculoskeletal, postural, and visual complaints.

However, the absence of a control group and the short duration of the intervention limit the ability to draw causal inferences. Future studies should employ randomized controlled designs, blinded assessments, longer follow-up periods, and integrated measures of engagement to more rigorously evaluate effectiveness and understand the psychological impact of hybrid virtual exercise programs.

LIMITATIONS

This study has several limitations that should be taken into account when interpreting the findings. First, the single-group pre–post design and absence of blinding limit causal inference. These constraints were largely due to the feasibility requirements of the TÜBİTAK 2209 student research framework, which provides restricted time and resources for implementing parallel control groups. Within this context, the study was intentionally structured as a preliminary feasibility investigation to explore the applicability, safety, and potential direction of intervention effects. Nevertheless, future research should incorporate randomized controlled designs with blinded assessors and larger samples to strengthen internal validity.

Second, the interpretation of ROSA scores warrants caution. The intervention focused on sensorimotor, oculomotor, and postural awareness, rather than workstation ergonomics, and no physical adjustments were made to chairs, screens, or desks. As such, stability in ROSA scores was expected and aligns with previous studies, which indicate that ROSA primarily captures static, equipment-related ergonomic parameters rather than behavioral or perceptual changes[6]. Additionally, the young, non-desk-based nature of the sample limits the relevance of workstation-based ergonomic measures, further reducing the likelihood of detecting short-term ergonomic changes.

Third, the homogeneity of the sample limits generalizability. Participants were predominantly young university students with similar academic routines and relatively low cumulative ergonomic load. While this homogeneity helped reduce confounding factors and improve internal consistency for a pilot study, it limits extrapolation to older adults or professional desk workers. Future investigations should include more diverse populations with varying ergonomic exposures.

Fourth, adherence and engagement were self-reported, which introduces potential bias. Although participation was encouraged through supervised sessions and daily video guidance, objective measurement tools such as engagement trackers or wearable sensors were not utilized. Reliance on self-reported symptoms, including headaches or eye discomfort, may also have introduced minor confounding effects.

Finally, the relatively short duration and moderate intensity of the program may not have been sufficient to elicit sustained psychological or neurophysiological adaptations, particularly regarding anxiety outcomes. Extending intervention length, refining task difficulty progression, and integrating real-time feedback may enhance both psychological and physical outcomes in future programs.

AUTHORS' CONTRIBUTIONS

AA: Conceptualization, Data curation, Formal Analysis, Writing – original draft, Writing – review & editing. **HNM:** Conceptualization, Data curation, Writing – original draft, Writing – review & editing. **SA:** Conceptualization, Data curation, Writing – original draft, Writing – review & editing. **AK:** Conceptualization, Data curation, Writing – original draft, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request, subject to ethical approval and institutional regulations.

REFERENCES

1. Almutairi H, Alhammad L, Aldossari B, Alonazi A. Prevalence and interrelationships of screen time, visual disorders, and neck pain among university students: a cross-sectional study at Majmaah University. Healthcare (Basel). 2024;12(20):2067. https://doi.org/10.3390/healthcare12202067
2. Emam MA, Hortobágyi T, Horváth AA, Ragab S, Ramadan M. Proprioceptive training improves postural stability and reduces pain in cervicogenic headache patients: a randomized clinical trial. J Clin Med. 2024;13(22):6777. https://doi.org/10.3390/jcm13226777
3. Kannappan SR, Alva J, Dsa VB. Examining the impact of the exercise package on eye and neck strain (EPENS) among adolescents: a pre-experimental study at a selected preuniversity college in Mangalore. J Educ Health Promot. 2024;13:459. https://doi.org/10.4103/jehp.jehp_20_24
4. Kaur K, Gurnani B, Nayak S, Deori N, Kaur S, Jethani J, et al. Digital eye strain- a comprehensive review. Ophthalmol Ther. 2022;11(5):1655-80. https://doi.org/10.1007/s40123-022-00540-9
5. Şambel Aykutlu M, Aykutlu HC, Özveren M, Garip R. Digital media use and its effects on digital eye strain and sleep quality in adolescents: a new emerging epidemic? PLoS One. 2024;19(12):e0314390. https://doi.org/10.1371/journal.pone.0314390
6. Barros FC, Moriguchi CS, Chaves TC, Andrews DM, Sonne M, Oliveira Sato T. Usefulness of the rapid office strain assessment (ROSA) tool in detecting differences before and after an ergonomics intervention. BMC Musculoskelet Disord. 2022;23(1):526. https://doi.org/10.1186/s12891-022-05490-8
7. Fischer-Grote L, Fössing V, Aigner M, Fehrmann E, Boeckle M. Effectiveness of online and remote interventions for mental health in children, adolescents, and young adults after the onset of the COVID-19 pandemic: systematic review and meta-analysis. JMIR Ment Health. 2024;11:e46637. https://doi.org/10.2196/46637
8. Bacık Yaman G, Hocaoglu C. Zoom yorgunluğu: bir gözden geçirme. Psikiyatr Güncel Yaklaşımlar. 2022;14(3):407-13. https://doi.org/10.18863/pgy.1032058
9. Weiniger SP, Schilaty ND. Interoceptive posture awareness and accuracy: a novel photographic strategy towards making posture actionable. Front Neurosci. 2024;18:1359594. https://doi.org/10.3389/fnins.2024.1359594
10. Khalsa SS, Adolphs R, Cameron OG, Critchley HD, Davenport PW, Feinstein JS, et al. Interoception and mental health: a roadmap. Biol Psychiatry Cogn Neurosci Neuroimaging. 2018;3(6):501-13. https://doi.org/10.1016/j.bpsc.2017.12.004
11. Eom T, Im S, Lee EH, Kim RJ, Ihm J. Effectiveness of virtual reality-based real-time ergonomics training on dental posture improvement. Int Dent J. 2025;75(6):103908. https://doi.org/10.1016/j.identj.2025.103908
12. Beyea D, Lim C, Lover A, Foxman M, Ratan R, Leith A, et al. Zoom fatigue in review: a meta-analytical examination of videoconferencing fatigue's antecedents. Comput Hum Behav Rep. 2025;17:100571. https://doi.org/10.1016/j.chbr.2024.100571
13. Koca S. Online eğitimin göz yüzeyine etkisinin değerlendirilmesi. MN Oftalmoloji. 2022;4(29):222-6.
14. Bayar B, Güp AA, Özen Oruk D, İpek Dongaz Ö, Doğu E, Bayar K. Development of the postural habits and awareness scale: a reliability and validity study. Int J Occup Saf Ergon. 2023;29(2):815-20. https://doi.org/10.1080/10803548.2022.2082694
15. Erdinc O. Upper extremity musculoskeletal discomfort among occupational notebook personal computer users: work interference, associations with risk factors and the use of notebook computer stand and docking station. Work. 2011;39(4):455-63. https://doi.org/10.3233/WOR-2011-1195
16. Sonne M, Villalta DL, Andrews DM. Development and evaluation of an office ergonomic risk checklist: ROSA--rapid office strain assessment. Appl Ergon. 2012;43(1):98-108. https://doi.org/10.1016/j.apergo.2011.03.008
17. Spielberg CHAD. Development of the STAI trait anxiety inventory. Interam J Psychol. 1971;3-4.
18. Zangirolami-Raimundo J, Raimundo RD, Silva Noll PRE, Santos WS, Leone C, Baracat EC, et al. Postmenopausal women's cognitive function and performance of virtual reality tasks. Climacteric. 2023;26(5):445-54. https://doi.org/10.1080/13697137.2023.2190511
19. Redivo VS, Olivier B. Time to re-think our strategy with musculoskeletal disorders and workstation ergonomics. S Afr J Physiother. 2021;77(1):1490. https://doi.org/10.4102/sajp.v77i1.1490
20. Zangirolami-Raimundo J, Noll PRES, Raimundo RD, Gonçalves GL, Urso EME, Bech GD, et al. Use of interventions involving virtual reality tasks during the climacteric: a systematic review. Climacteric. 2022;25(6):543-51. https://doi.org/10.1080/13697137.2022.2088275
21. Silva-Jose C, Nagpal TS, Coterón J, Barakat R, Mottola MF. The 'new normal' includes online prenatal exercise: exploring pregnant women's experiences during the pandemic and the role of virtual group fitness on maternal mental health. BMC Pregnancy Childbirth. 2022;22(1):251. https://doi.org/10.1186/s12884-022-04587-1
22. Canonica AC, Alonso AC, Brech GC, Peterson M, Luna NMS, Busse AL, et al. Adaptation to the driving simulator and prediction of the braking time performance, with and without distraction, in older adults and middle-aged adults. Clinics (Sao Paulo). 2023;78:100168. https://doi.org/10.1016/j.clinsp.2023.100168

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251434

Comparison of different artificial intelligence tools' answers to questions related to early intervention: ChatGPT versus Gemini

Ayşegül Yildirim[1], Aslıhan Turan[2], Hasan Gerçek[3]*, Gülhan Çaka[4]

SUMMARY

OBJECTIVE: The aim of this study was to assess the quality and readability of ChatGPT and Gemini's responses to frequently asked questions about early intervention for individuals with at-risk infants.

METHODS: Ten frequently asked questions about early intervention were selected by three researchers (a child development specialist, a physiotherapist, and a midwife) from a list generated by ChatGPT and Gemini. Questions were sent to ChatGPT version 4.0 and Gemini 1.5, and initial responses were recorded without follow-up queries. Ten independent experts (two special education specialists, two child development specialists, two physiotherapists, two midwives, and two pediatricians) The quality of ChatGPT and Gemini's responses was assessed using a four-grade rating system. Readability levels were analyzed using the Flesch-Kincaid Grade Level through WordCalc software.

RESULTS: One of the answers given by ChatGPT was of higher quality than Gemini (p=0.025), while one answer given by Gemini was of higher quality than ChatGPT (p=0.033). The answers to the other questions were of similar quality, with Gemini having a lower level.

CONCLUSION: This study compares the quality and readability of the answers given by artificial intelligence-based language models to demonstrate their potential to appeal to different user groups. While the models generally provided answers of similar quality, quantitative differences in readability were observed, suggesting potential suitability for different audiences. These findings contribute to understanding the role of AI tools in health communication.

KEYWORDS: Artificial intelligence. Early intervention. Health communication. Readability.

INTRODUCTION

Artificial intelligence technology has experienced very rapid developments in recent years, and this has become a part of daily life by affecting studies in communication, entertainment, education, health, and many other fields[1,2]. These rapid developments in the field of artificial intelligence (AI) have led to different AI models[3]. Large language models (LLMs), as AI models, are AI systems trained on countless words generated from books, articles, and internet-based content[4].

The most popular is ChatGPT, a language modeling tool developed by OpenAI that allows people to chat with a machine on different topics, generating responses to user-specified prompts[5,6]. ChatGPT is a generative AI chatbot, and other AI models are being developed through similar processes[4].

Gemini is another powerful language model. Recognizing the power and impact of AI, Google DeepMind has developed Gemini, an intelligent chat tool with advanced capabilities. Gemini gives users access to accurate and relevant answers by utilizing the database accumulated by Google[7]. Gemini is trained not only on text data but also on various types of data, such as audio, image, and video, so it works by processing integrated data types[8].

ChatGPT is leading the way in tailoring personalized mental health support and improving chronic condition management. Gemini advanced offers support in facilitating early detection of diseases and medical decision-making[9]. People can use AI models when searching for health-related information[10]. Parents use online resources to learn about their baby's development and health[11,12].

At-risk infants are characterized by adverse environmental and biological factors that occur during pregnancy, at birth, or after birth and contribute to the development of neurodevelopmental disorders and increased mortality[13]. Parents who have at-risk babies may have problems in terms of travelling

[1]KTO Karatay University, Vocational School of Health Sciences, Department of Child Development – Konya, Turkey.
[2]KTO Karatay University, Faculty of Health Sciences, Department of Midwifery – Konya, Turkey.
[3]Acıbadem University, Faculty of Health Sciences, Department of Physical Therapy and Rehabilitation, Istanbul, Turkey.
[4]Mersin Bozyazı State Hospital, Department of Pediatric – Mersin, Turkey.
*Corresponding author: hasangercek42@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on January 01, 2026. Accepted on February 14, 2026.
Scientific Editor: José Maria Soares Júnior

to the health institution, transportation, cost, time, etc. With the development of AI models, it can be a good support for parents to guide and help their children[14].

According to a study conducted at the University of Kansas, parents seeking information about their children's health trusted AI more than health professionals, and evaluated a text generated by AI as reliable and credible[15]. Given the rapid development and increasing popularity of AI and its impact on the dissemination of health-related information, it is extremely important to assess the accuracy and quality of ChatGPT and Gemini[16]. To the best of our knowledge, no study has investigated the questions of parents with at-risk infants in AI models. The aim of this study is to compare the quality and readability of ChatGPT and Gemini's answers to the questions asked by individuals with high-risk babies about early intervention.

METHODS

Question collection process

The purpose of using ChatGPT and Gemini to generate an initial pool of questions was not to identify empirically validated "frequently asked questions," but rather to capture how LLMs conceptualize common parental concerns regarding early intervention. This approach was chosen deliberately to align with the study's primary aim: To compare the responses generated by AI systems to questions formulated within their own knowledge frameworks. Therefore, the generated questions reflect the models' internal representations of commonly asked questions rather than real-world parental query frequencies.

From the initial pool of 100 questions (50 generated by ChatGPT and 50 by Gemini), a final set of 10 questions was selected through a structured consensus process conducted by three researchers (a child development specialist, a physiotherapist, and a midwife). The selection process consisted of the following steps: (a) Relevance screening: Questions not directly related to early intervention in at-risk infants were excluded. (b) Overlap and redundancy assessment: Conceptually similar or duplicate questions were merged or eliminated. (c) Clarity and comprehensibility: Questions with ambiguous wording or unclear clinical intent were excluded. (d) Multidisciplinary relevance: Priority was given to questions addressing medical, developmental, and rehabilitative aspects of early intervention. All three researchers independently reviewed the questions and subsequently met to reach consensus through discussion. Only questions approved unanimously were included in the final analysis (Figure 1 and Table 1).

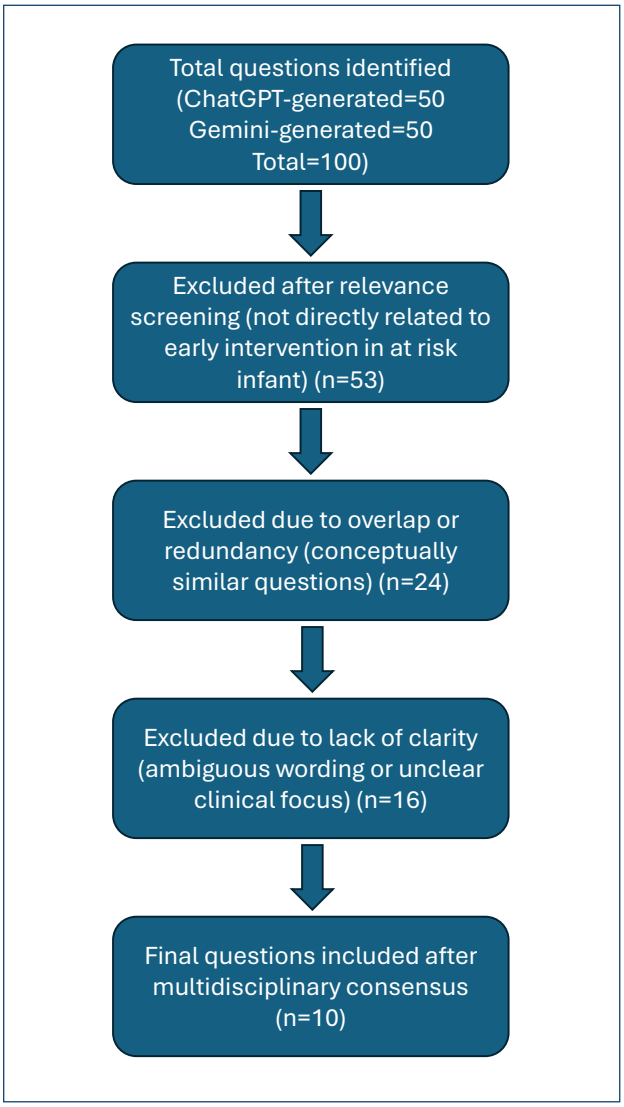


Figure 1. Flowchart illustrating the selection process of the final 10 questions.

Table 1. Questions.

No	Question
1	What exactly is early intervention?
2	How long will the early intervention program last?
3	Which therapies will be used in the early intervention program?
4	In which cases is early intervention recommended?
5	Who is a member of the early intervention team?
6	How often should we go to therapy?
7	What arrangements should we make for early intervention at home?
8	Is early intervention always successful?
9	What should we do if we don't get results?
10	What are the consequences of late initiation of early intervention?

ChatGPT and Gemini usage

The questions were sent using ChatGPT version 4.0 and Gemini version 1.5. In order to prevent the applications from being affected by past searches, the browser history was cleared first, then a new ChatGPT and Gemini account was opened for the first time. Then, 10 questions determined by the researchers were asked in order. For each question, only a single, isolated prompt was used, and the first response generated by the applications was recorded without follow-up queries or iterative refinement. This approach was intentionally chosen to standardize comparisons; however, it does not reflect real-world user behavior, where individuals often engage in iterative questioning or prompt refinement when seeking health-related information.

Assessment

The quality of the responses was assessed by 10 independent investigators (2 special education specialists, 2 child development specialists, 2 physiotherapists, 2 midwives, and 2 pediatricians) using the four-grade system described by Mika et al.[17]. The scale is scored from 1 to 4. One means a perfect answer, and four means an inadequate answer (Supplementary Table 1).

The readability of the apps' answers was assessed using WordCalc software, where the answers to each question were pasted into a readability calculator and the corresponding Flesch-Kincaid Grade Level was recorded. The evaluators assessed the answers of both ChatGPT and Gemini apps, blind to the app that provided the answers.

Statistic

The software package Statistical Package for the Social Sciences Statistics for Windows, Version 29.0 (IBM Corp., Armonk, NY, USA) was used to analyze the data. The results of the quality assessment of the answers are given as median and interquartile range (IQR). The Wilcoxon signed-rank test was used to compare paired quality scores between ChatGPT and Gemini for each question, as the data were ordinal in nature and derived from a four-point rating scale. This non-parametric test was considered appropriate because the assumptions of normality cannot be met for ordinal scale data.

RESULTS

The comparison of quality scores between ChatGPT and Gemini is presented in Table 2. ChatGPT provided significantly higher-quality responses for Question 4 (p=0.025), whereas Gemini provided higher-quality responses for Question 7 (p=0.033). No statistically significant differences were observed for the remaining questions (Table 2).

ChatGPT demonstrated lower Flesch-Kincaid Grade Level scores than Gemini for eight out of ten questions, while Gemini demonstrated lower scores for two questions (Questions 1 and 4). Detailed readability scores for each question are presented in Table 3.

DISCUSSION

In this study, we evaluated the quality and readability of the ChatGPT version 4.0 and the Gemini version 1.5 application's responses about early intervention from families with at-risk infants. Our results showed that the responses of ChatGPT version 4.0 and the Gemini 1.5 app had similar quality. The observation that one model outperformed the other on isolated questions should be interpreted cautiously. Given that statistically significant differences were observed for only two out

Table 2. Comparison of the answers given by the applications.

No	Quality of answer				z	p
	ChatGPT		Gemini			
	Median	IQR	Median	IQR		
1	1.50	1.25	2.00	0.25	0.447	0.655
2	2.00	1.25	2.00	1.00	1.667	0.096
3	2.00	1.25	2.00	1.25	0.477	0.655
4	1.50	1.00	2.00	0.50	2.236	**0.025***
5	1.00	1.00	2.00	1.25	1.414	0.157
6	1.50	1.25	1.50	1.00	0.272	0.785
7	2.00	3.25	1.00	1.00	2.126	**0.033***
8	1.00	1.00	1.00	1.00	0.000	>0.999
9	1.00	1.00	2.00	1.00	1.732	0.083
10	2.00	0.00	2.00	0.25	1.000	0.317

IQR: interquartile range; Z: Wilcoxon signed test. *p<0.05. Bold values indicate statistically significant results.

Table 3. Flesch-Kincaid Degree Level of questions.

Question	ChatGPT	Gemini
1	11.2	10
2	9.8	12
3	11.9	12
4	12	11.8
5	9.2	12.2
6	11.3	11.9
7	9.1	9.5
8	11.9	12
9	10.8	11
10	11.3	12.1

of ten questions, these findings may reflect question-specific variation or inherent differences in model training rather than systematic superiority of one model over the other.

Gemini was reported to be more scientifically interesting and accurate in spelling than ChatGPT. Comparison of the answers given by the applications[18]. In the evaluation of the recommendations given by ChatGPT and Gemini applications, ChatGPT and Gemini showed similar performance in pediatric orthopedic conditions[19], while for family education on retinopathy of prematurity, Gemini was reported to have better readability scores but lower reliability scores[20]. Among the ChatGPT (GPT-4), Google Bard (Gemini Pro), Microsoft Bing Chat, and Google SGE applications, ChatGPT showed an overall high performance in providing valuable and reliable information to carers of children with cancer[21] and it was found that ChatGPT had higher readability[22] than Gemini, but the quality of Gemini was higher. ChatGPT and Gemini both demonstrate impressive capabilities in text comprehension and generation, and each has distinct strengths in several areas[23]. Studies comparing ChatGPT and Gemini in different fields are available in the literature, but there are no studies evaluating the accuracy of ChatGPT and Gemini responses in mothers with high-risk infants. The results of the study are generally similar to the literature. Therefore, the quality differences observed in this study should not be interpreted as overall model dominance, but rather as context-dependent performance variations.

Differences observed between comparative studies of LLMs may be partly explained by the clinical domain in which the models are evaluated. In procedural and guideline-driven specialties such as orthopedics, higher-quality responses may be associated with structured, protocol-based knowledge, whereas in ophthalmology or neonatal care, the need for precise medical terminology may increase the perceived technicality and reading level of responses. In contrast, early intervention focuses heavily on family-centered communication, developmental guidance, and interdisciplinary coordination, which may favor more explanatory and accessible language rather than strictly technical detail. Taken together, variations across comparative studies likely arise from differences in clinical complexity, communication goals, and audience expectations rather than from consistent model-specific advantages.

CONCLUSION

This study assessed the quality and readability of responses provided by ChatGPT version 4.0 and GEMİNİ application to questions related to early intervention for individuals with at-risk infants. The findings showed that both ChatGPT and Gemini produced responses of similar quality, however, Gemini generally required a higher reading level than ChatGPT.

While both applications provided acceptable answers, in terms of readability, ChatGPT responses tended to demonstrate lower reading grade levels compared to Gemini, suggesting potentially greater accessibility for general users. However, the observed differences were modest and varied across individual questions. Future research should broaden the scope by including different user demographics and investigating the accuracy of responses on other healthcare topics. Furthermore, integrating user feedback can enhance the personalized utility and practical application of these AI systems in healthcare communication. Future studies should incorporate iterative prompt strategies to better approximate real-world user–AI interactions in health information seeking.

Limitations

This study is the first to compare the quality and readability of responses of two advanced AI systems, ChatGPT-4.0 and Gemini, in the context of early intervention for at-risk infants. This study provides novel insights into the performance of LLMs in early intervention contexts. However, this study had some limitations. A major limitation of this study is that the initial question pool was generated by LLMs rather than derived from empirically collected parental queries. Consequently, the selected questions represent the models' perception of common concerns rather than actual question frequencies among parents. This methodological choice may introduce bias and limit the generalizability of the findings to real-world parental information-seeking behavior. Another important limitation of this study is the use of single, non-iterative prompts for each question. In real-world settings, users frequently refine their questions, request clarifications, or ask follow-up questions when interacting with AI-based systems. The absence of iterative prompting in this study may limit the generalizability of the findings, as the quality and readability of AI-generated responses may improve or change with prompt engineering or continued interaction. The study only analyzed a specific set of questions related to early intervention for at-risk infants. A larger pool of questions could be examined to generalize the findings. Another limitation is that the study evaluated AI responses in a controlled environment and did not include real user feedback.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this study, the authors used ChatGPT Open AI and Gemini to identify questions and receive responses to the questions asked. After using this tool/service, the authors revised the content as necessary, but the responses were not

edited as the purpose of this study was to evaluate the effectiveness of ChatGPT and Gemini in the first place, as described in the article. The authors take full responsibility for the content of the publication.

AUTHORS' CONTRIBUTIONS

AY: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Writing – original draft. **AT:** Data curation, Investigation, Validation, Writing – review & editing. **HG:** Conceptualization, Formal Analysis, Methodology, Software, Supervision, Validation, Visualization, Writing – review & editing. **GÇ:** Investigation, Resources, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Talati D, Joe B, Smart G. AI (artificial intelligence) in daily life. Authorea Preprints. 2024.
2. Vinothkumar J, Karunamurthy A. Recent advancements in artificial intelligence technology: trends and implications. Int J Multidiscip Sci Res Dev. 2023;2(1):1-11.
3. Pugachev AA, Kharchenko AV, Sleptsov NA. Transforming the future: a review of artificial intelligence models. RUDN J Stud Literature Journalism. 2023;28(2):355-67.
4. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023;29(8):1930-40. https://doi.org/10.1038/s41591-023-02448-8
5. Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. JMIR Med Educ. 2023;9:e46885. https://doi.org/10.2196/46885
6. Reynolds K, Tejasvi T. Potential use of ChatGPT in responding to patient questions and creating patient resources. JMIR Dermatol. 2024;7:e48451. https://doi.org/10.2196/48451
7. Saeidnia HR. Welcome to the Gemini era: Google DeepMind and the information industry. Libr Hi Tech News. 2023;43(3). https://doi.org/10.1108/LHTN-12-2023-0214
8. Perera P, Lankathilake MA. Preparing to revolutionize education with the multi-model GenAI tool Google Gemini? A journey towards effective policy making. J Adv Educ Philos. 2023;7(08):246-53. https://doi.org/10.36348/jaep.2023.v07i08.001
9. Alhur A. Redefining healthcare with artificial intelligence (AI): the contributions of ChatGPT, Gemini, and co-pilot. Cureus. 2024;16(4):e57795. https://doi.org/10.7759/cureus.57795
10. Kassab J, Hadi Hajjar A, Wardrop RM, Brateanu A. Accuracy of online artificial intelligence models in primary care settings. Am J Prev Med. 2024;66(6):1054-9. https://doi.org/10.1016/j.amepre.2024.02.006
11. Coughler C, M Burke S, Cardy JO. Analysis of the quality of online resources for parents of children who are late to talk. Autism Dev Lang Impair. 2020;5:2396941520917940. https://doi.org/10.1177/2396941520917940
12. Orlando JM, Cunha AB, Alghamdi ZS, Lobo MA. How do parents and early intervention professionals utilize educational resources about infant development and play? Early Hum Dev. 2023;180:105763. https://doi.org/10.1016/j.earlhumdev.2023.105763
13. Gercek H, Balci NC, Unuvar BS, Kemers SN, Demirsoz M, Cakir D, et al. Turkish adaptation and psychometric properties of the standardized infant neurodevelopmental assessment neurological scale in Turkish at-risk infants. IJ Pediatrics. 2024;34(4):e142447. https://doi.org/10.5812/ijp-142447
14. Ashraf M. A systematic review on the potential of AI and ChatGPT for parental support and child well-being. arXiv preprint arXiv:240709492. 2024. https://doi.org/10.48550/arXiv.2407.09492
15. Leslie-Miller CJ, Simon SL, Dean K, Mokhallati N, Cushing CC. The critical need for expert oversight of ChatGPT: prompt engineering for safeguarding child healthcare information. J Pediatr Psychol. 2024;49(11):812-7. https://doi.org/10.1093/jpepsy/jsae075
16. Lee TJ, Campbell DJ, Patel S, Hossain A, Radfar N, Siddiqui E, et al. Unlocking health literacy: the ultimate guide to hypertension education from ChatGPT versus Google Gemini. Cureus. 2024;16(5):e59898. https://doi.org/10.7759/cureus.59898
17. Mika AP, Martin JR, Engstrom SM, Polkowski GG, Wilson JM. Assessing ChatGPT responses to common patient questions regarding total hip arthroplasty. J Bone Joint Surg Am. 2023;105(19):1519-26. https://doi.org/10.2106/JBJS.23.00209
18. AlSagri HS, Farhat F, Sohail SS, Saudagar AKJ. ChatGPT or Gemini: who makes the better scientific writing assistant? J Acad Ethics. 2024:23:1121-35.
19. Pirkle S, Yang J, Blumberg TJ. Do ChatGPT and Gemini provide appropriate recommendations for pediatric orthopaedic conditions? J Pediatr Orthop. 2025;45(1):e66-71. https://doi.org/10.1097/BPO.0000000000002797
20. Durmaz Engin C, Karatas E, Ozturk T. Exploring the role of ChatGPT-4, BingAI, and Gemini as virtual consultants to educate families about retinopathy of prematurity. Children (Basel). 2024;11(6):750. https://doi.org/10.3390/children11060750
21. Sezgin E, Jackson DI, Kocaballi AB, Bibart M, Zupanec S, Landier W, et al. Can large language models aid caregivers of pediatric cancer patients in information seeking? A cross-sectional investigation. Cancer Med. 2025;14(1):e70554. https://doi.org/10.1002/cam4.70554
22. Hancı V, Ergün B, Gül Ş, Uzun Ö, Erdemir İ, Hancı FB. Assessment of readability, reliability, and quality of ChatGPT®, BARD®, Gemini®, Copilot®, Perplexity® responses on palliative care. Medicine (Baltimore). 2024;103(33):e39305. https://doi.org/10.1097/MD.0000000000039305
23. Rane N, Choudhary S, Rane J. Gemini versus ChatGPT: applications, performance, architecture, capabilities, and implementation. J Appl Artif Intel. 2024;5(1):69-93

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251460

The effect of sodium-glucose cotransporter-2 inhibitors on haemoglobin and haematocrit in diabetic kidney disease

Süleyman Karaköse[1]*, Beyza Doğan[1], Fatih Ergül[1], İbrahim Güney[1]

SUMMARY

OBJECTIVE: The aim of this study was to evaluate the impact of sodium-glucose cotransporter-2 inhibitors on haematological parameters and other clinical markers in this patient population.

METHODS: We retrospectively analyzed the medical records of 150 diabetic patients with chronic kidney disease who initiated sodium-glucose cotransporter-2 inhibitor therapy, comparing their baseline and 3-month post-treatment laboratory and clinical data.

RESULTS: Following 3 months of treatment with sodium-glucose cotransporter-2 inhibitors, mean haemoglobin and haematocrit levels significantly increased, even in patients with an estimated glomerular filtration rate below 60 mL/min/1.73 m^2. This therapy also led to significant reductions in the spot urine albumin-to-creatinine ratio, as well as in both systolic and diastolic blood pressure.

CONCLUSION: Sodium-glucose cotransporter-2 inhibitor therapy was associated with significant increases in haemoglobin and haematocrit levels in diabetic patients with chronic kidney disease. These findings suggest that sodium-glucose cotransporter-2 inhibitors may represent a potential area requiring further investigation for the treatment of anaemia associated with diabetic chronic kidney disease.

KEYWORDS: Anaemia. Dapagliflozin. Empagliflozin. Chronic kidney disease.

INTRODUCTION

Diabetic kidney disease (DKD) is the leading cause of chronic kidney disease (CKD) globally. The presence of CKD in diabetic patients is a significant independent risk factor for both cardiovascular disease (CVD) and increased mortality[1].

Current data show that anaemia is associated with an increased risk of CKD progression, CVD, and mortality. While multiple factors contribute to anaemia as the estimated glomerular filtration rate (eGFR) declines, the primary cause is the impaired secretion of erythropoietin (EPO) due to the loss of functional nephrons[2].

Sodium-glucose cotransporter-2 (SGLT2) inhibitors regulate blood glucose by promoting renal glucose excretion. Beyond their glucose-lowering effects, they also demonstrate pleiotropic actions that prevent the progression of renal disease and heart failure. Consequently, SGLT2 inhibitors have transitioned from being merely hypoglycaemic agents to being recognized as essential cardiorenal risk-reducing agents in patients with DKD[3].

While the increase in haemoglobin (Hb) and haematocrit (Hct) from SGLT2 inhibitor therapy is partly attributed to haemoconcentration, this mechanism alone seems insufficient. It is hypothesized that additional factors, such as a positive effect on haematopoiesis, also contribute to this rise[4,5]. Beyond their diuretic effects, SGLT2 inhibitors may influence renal hypoxia and haematopoiesis by increasing Hct levels[5,6].

A study involving 4,304 participants demonstrated that dapagliflozin, an SGLT2 inhibitor, was associated with an increase in Hct, the correction of anaemia, and a reduced risk of incident anaemia in patients with CKD, irrespective of the presence of type 2 diabetes mellitus (T2DM)[7].

A study evaluating the effect of GLP-1 receptor agonists and SGLT2 inhibitors on anaemia, involving 13,799 participants, demonstrated that SGLT2 inhibitors reduced the incidence of anaemia[8].

Data on their impact on Hb and Hct are limited, particularly in patients with DKD. Therefore, this retrospective study was conducted to specifically evaluate the effect of SGLT2 inhibitors on Hb and Hct levels in this patient population.

[1]University of Health Science, Konya City Hospital, Department of Nephrology – Konya, Turkey.
*Corresponding author: suleymankarakose@yahoo.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on August 13, 2025. Accepted on January 31, 2026.
Scientific Editor: José Maria Soares Júnior

METHODS

Study design and population

This was a single-centre retrospective cohort study conducted at the nephrology clinic. The study protocol was approved by the ethics committee (approval no: 2024/029) and was conducted following the Declaration of Helsinki.

The study population included patients followed at our nephrology outpatient clinic between October 2023 and March 2024. We retrospectively reviewed the files of patients with Stage 1–3 CKD who had initiated treatment with either dapagliflozin 10 mg or empagliflozin 10 or 25 mg. The primary mechanism of SGLT2 inhibition, involving the reduction of glomerular hyperfiltration, is largely established at both dose levels. Furthermore, our initial analysis revealed that when separated, the cohort taking empagliflozin 10 mg and the cohort taking empagliflozin 25 mg had insufficient statistical power to detect meaningful differences in the primary and secondary endpoints (glycated haemoglobin [%HbA1c], eGFR, and urine albumin-to-creatinine ratio [UACR]). Combining the groups enhanced the n (total sample size for the empagliflozin arm) and mitigated the risk of Type II errors (false negatives). The CKD stages were determined based on KDIGO guidelines.

Baseline and 3-month clinical and biochemical data were collected from patient files. The recorded laboratory results included Hb, Hct, mean corpuscular volume (MCV), potassium (K), creatinine, albumin, C-reactive protein (CRP), uric acid, glucose, %HbA1c, and spot UACR. Blood pressure (BP) measurements were recorded following standard guidelines.

From an initial pool of 244 patients, a final cohort of 150 patients was included after applying strict exclusion criteria. Patients were excluded if they had a history of malignancy, gastrointestinal bleeding, kidney transplantation, autoimmune disease, valvular disease, heart failure, active infection, or a deficiency in iron, folic acid, or vitamin B12.

In our retrospective cohort of patients with CKD stages 1, 2, and 3, the dataset did not include cases of iron deficiency anaemia, iron supplementation, or ESA treatment at baseline or during follow-up.

Biochemical and clinical data were collected from each patient's files at two time points: baseline and 3 months after treatment. The recorded laboratory results included Hb, Hct, and MCV; potassium (K), creatinine, albumin, and uric acid; CRP; glucose and %HbA1c; and spot UACR. Additionally, BP measurements were routinely recorded at each patient visit, following standard clinical guidelines[9].

Statistical analysis

Statistical analyses were performed using SPSS version 22.0 (Statistical Package for Social Sciences). The Kolmogorov-Smirnov test was used to assess for normality. Data are presented as mean±standard deviation for normally distributed variables and as median (minimum–maximum) for skewed variables. For pairwise comparisons, the paired-samples t-test was used for normally distributed data, while the Wilcoxon signed-rank test was performed for non-normally distributed data. $p<0.05$ was considered statistically significant.

RESULTS

A total of 150 patients were included in the study, with 101 patients (67.3%) receiving dapagliflozin and 49 patients (32.7%) receiving empagliflozin. The mean age of all patients was 60.23±10.2 years, with an age range of 27–82 years.

At the beginning of the SGLT2 inhibitor treatment, the mean Hb and Hct levels were statistically significantly lower than at the third month (13.54±1.74 g/dL vs. 14.03±1.77 g/dL, $p<0.001$; and 41.09±5.04% vs. 42.99±4.63%, $p<0.001$, respectively). White blood cell (WBC) counts were not statistically different. Fasting glucose levels did not show a statistically significant difference, but HbA1c values were statistically lower at the third month (7.6 vs. 7.5, $p<0.001$). The spot UACR was also significantly lower at the third month (204.5 vs. 81, $p<0.001$). Additionally, there was a decrease in systolic blood pressure (SBP) and diastolic blood pressure (DBP) levels with statistically significant differences at the third month (147 mmHg vs. 135 mmHg, $p<0.001$; and 80 mmHg vs. 76 mmHg, $p<0.001$). As expected, the creatinine value was lower (1.1 mg/dL [0.49–2.14] vs. 1.20 mg/dL [0.55–2.62]) and eGFR value was higher (60.5 [30–130] mL/dk/1.73 m^2 vs. 58.5 [21–133] mL/dk/1.73 m^2) at baseline when compared to the third month's values with statistically significant differences ($p<0.001$ and $p<0.001$, respectively) (Table 1).

Hb and Hct levels increased significantly in both the dapagliflozin and empagliflozin groups. HbA1c levels were lower at the third month for patients on dapagliflozin 10 mg (7.92±1.85 vs. 7.61±1.48, $p=0.007$) and empagliflozin 25 mg (7.7 vs. 7.7, $p=0.010$). The SBP levels were statistically lower at the third month in all patient groups (Table 2).

In the subgroup of patients with baseline eGFR between 25 and 60 mL/min/1.73 m^2, Hb and Hct values increased significantly in both dapagliflozin (13.15±1.57 vs. 13.49±1.72, $p=0.010$; and 40.31±4.53 vs. 41.60±4.48, $p=0.001$) and empagliflozin groups (13.12±1.55 vs. 13.73±1.67, $p<0.001$; and 39.77±3.99 vs. 41.94±4.50, $p<0.001$). SBP levels were

Table 1. The laboratory and clinical results of all patients.

Parameters	Baseline (n=150)	At the third month (n=150)	p
Cre (mg/dL)	1.1 (0.49–2.14)	1.20 (0.55–2.62)	*<0.001*
eGFR (mL/min)	60.5 (30–130)	58.5 (21–133)	*<0.001*
K^+ (mmol/l)	4.8 (3.2–6.0)	4.6 (3.5–5.8)	0.06
Hb (g/dL)	13.54±1.74	14.03±1.77	*<0.001*
Hct	41.09±5.04	42.99±4.63	*<0.001*
MCV	86.66±4.71	85.95±5.54	0.028
WBC	8,325 (864–16,200)	8,065 (1,040–16,640)	0.656
Glucose (mg/dL)	143 (59–350)	141 (61–418)	0.358
HbA1c	7.6 (5–14.9)	7.5 (5–12.6)	*<0.001*
CRP	2.0 (0–31)	2.0 (0–99)	0.124
Albuminuria (mg/dL)	204.5 (0–5,480)	81 (0–6,259)	*<0.001*
SBP (mmHg)	147 (99–194)	135 (100–184)	*<0.001*
DBP (mmHg)	80 (60–121)	76 (50–109)	*<0.001*

Cre: creatinine; GFR: estimated glomerular filtration rate; K^+: potassium; MCV: mean corpuscular volume; Hb: haemoglobin; Hct: haematocrit; WBC: white blood cell; CRP: C-reactive protein; SBP: systolic blood pressure; DBP: diastolic blood pressure. Italic values indicate statistical significance (p<0.05).

also significantly lower at the third month in both subgroups (150.6±21.3 vs. 137.9±18.8, p<0.001 for dapagliflozin; and 149.6±20.9 vs. 138.2±20.9, p=0.002 for empagliflozin) (Table 3).

DISCUSSION

Our study found that SGLT2 inhibitor treatment is associated with a statistically significant increase in both Hb and Hct levels in patients with DKD. This effect was also observed in the subgroup of patients with an eGFR below 60 mL/min/1.73 m^2. The small but statistically significant increase observed in our study can be interpreted as suggestive evidence that the treatment is associated with activation of the haematopoietic pathway. This finding contributes to the existing body of literature by indicating a potential link between SGLT2 inhibitor use and modest haematological changes, which may be relevant to the discussion of managing anaemia, particularly in the context of progressive CKD. Due to the retrospective nature of our study, we were unable to measure routine follow-up EPO levels; consequently, we cannot definitively establish the contribution of SGLT2 inhibitors to EPO elevation, thus precluding strong causal inferences. However, despite this limitation, we believe our results are suggestive of a positive influence of SGLT2 inhibitors on anaemia-related parameters.

Furthermore, we observed significant reductions in the spot UACR, HbA1c levels, and SBP. Also, our study reported an initial increase in serum creatinine and subsequent decrease in eGFR, which reflect the expected and desired pharmacological effect of SGLT2 inhibitors and do not indicate a genuine worsening of renal function. SGLT2 inhibitors exert their effect by modulating the tone of the afferent arteriole and reducing glomerular hyperfiltration. This action leads to a transient and mild initial decline in eGFR (10–15%) due to changes in renal haemodynamics. These findings are consistent with the results of landmark trials such as DAPA-CKD, CREDENCE, and EMPA-KIDNEY, which have established the cardiorenal protective benefits of SGLT2 inhibitors[8-10]. Ultimately, our results underscore the importance of SGLT2 inhibitors in the comprehensive management of DKD, as controlling key risk factors such as hyperglycaemia, hypertension, proteinuria, and anaemia is critical for slowing disease progression[9,11,12]. According to current literature, the beneficial effects of SGLT2 inhibitors on DKD are primarily attributed to their actions on proteinuria, hypertension, and hyperglycaemia[13,14]. Although anaemia is a recognized factor in the progression of DKD, the effect of SGLT2 inhibitors on Hb and Hct levels in this specific patient population has not been fully elucidated. In light of this, our study aimed to specifically investigate the impact of SGLT2 inhibitors on these haematological parameters in patients with chronic DKD.

In different well-designed studies on patients with heart failure, an increase in Hb and Hct levels has been reported after SGLT2 inhibitor treatment[15-18]. At the first glance, this rise in Hct might be attributed solely to haemoconcentration caused by the diuretic effects of SGLT2 inhibitors. However, findings from the DAPA-HF study challenge this simple explanation. The study reported a consistent and similar degree of increase in Hct levels in patients both with and without diabetes mellitus.

Furthermore, the rise in Hct seemed to peak after approximately 4 months of treatment, which is a considerably longer duration than what would be expected from the diuretic effect alone. This suggests that additional mechanisms, beyond haemoconcentration, are likely contributing to these haematological changes[19]. Also, clinical studies have demonstrated that serum EPO levels increase within 4 weeks of initiating SGLT2 inhibitor therapy[20,21]. It is therefore hypothesized that

Table 2. The laboratory and clinical results of patients according to the drugs used.

	Dapagliflozin (n=101)	Empagliflozin 10/25 (n=49)	Empagliflozin 10 (n=23)	Empagliflozin 25 (n=26)
Cre				
Baseline	1.10 (0.56–2.14)	1.15±0.32	1.23±0.33	1.09±0.33
Third month	1.22 (0.55–2.62)	1.21±0.38	1.24±0.34	1.19±0.40
p	*<0.001*	*0.019*	0.748	*0.002*
eGFR				
Baseline	61 (30–118)	57 (30–130)	54 (30–130)	57.5 (34–109)
Third month	58 (21–113)	59 (26–133)	52 (30–133)	59.5 (29–104)
p	*<0.001*	0.114	0.917	0.290
K^+				
Baseline	4.75±0.48	4.71±0.47	4.50 (4.0–5.5)	4.95 (3.2–5.4)
Third month	4.61±0.49	4.66±0.45	4.4 (3.7–5.3)	4.90 (4.2–5.5)
p	*0.001*	0.390	0.300	0.406
Hb				
Baseline	13.51±1.68	13.66±1.84	13.31±1.92	13.96±1.76
Third month	13.96±1.71	14.20±1.87	13.87±1.98	14.50±1.75
p	*<0.001*	*<0.001*	*0.003*	*0.001*
Hct				
Baseline	40.94±5.25	41.41±4.60	40.94±4.79	41.83±4.47
Third month	42.84±4.52	43.30±4.88	42.46±5.24	44.04±4.51
p	*<0.001*	*<0.001*	*0.010*	*<0.001*
HbA1c				
Baseline	7.92±1.85	8.1±1.43	8.3 (5.6–11.3)	8.3 (6.0–10.8)
Third month	7.61±1.48	7.74±1.38	7.5 (5.6–12.6)	7.7 (6.1–9.9)
p	*0.007*	*0.007*	0.217	*0.010*
Albuminuria				
Baseline	259 (0–4,817)	80 (0–5,480)	51.5 (0–3,417)	102 (0–5,480)
Third month	111 (0–6,259)	10 (0–4,490)	3 (0–4,490)	39 (0–1,885)
p	*<0.001*	*0.005*	0.209	*0.005*
SBP				
Baseline	147.1±20.2	147.7±20.8	144.4±22.9	150.6±18.8
Third month	135.5±18.6	137.4±21.3	131.9±23.7	142.3±18.0
p	*<0.001*	*<0.001*	*<0.001*	*0.019*
DBP				
Baseline	81.4±10.6	80.5±8.1	78.7±5.9	82.1±9.4
Third month	76.7±10.3	76.3±9.1	73.8±7.5	78.5±0.9
p	*<0.001*	*0.003*	*0.003*	0.120

Cre: creatinine; GFR: estimated glomerular filtration rate; K^+: potassium; Hb: haemoglobin; Hct: haematocrit; SBP: systolic blood pressure; DBP: diastolic blood pressure. Italic values indicate statistical significance ($p<0.05$).

Table 3. The laboratory results of the patients who had 25<eGFR<60 mL/dk/1.73 m^2.

Parameters	Baseline	At the third month	p
Cre			
Dapagliflozin (n=45)	1.42±0.29	1.53±0.41	*0.004*
Empagliflozin 10/25 mg (n=27)	1.40 (1.0–1.78)	1.54 (0.9–1.91)	*0.046*
eGFR			
Dapagliflozin (n=45)	47 (30–58)	41 (21–83)	*0.008*
Empagliflozin 10/25 mg (n=27)	50 (30–58)	43 (29–82)	0.118
Potassium			
Dapagliflozin (n=45)	4.88±0.53	4.66±0.58	*0.003*
Empagliflozin 10/25 mg (n=27)	4.7 (3.2–5.4)	4.66 (3.7–5.5)	0.198
Hb			
Dapagliflozin (n=45)	13.15±1.57	13.49±1.72	*0.010*
Empagliflozin 10/25 mg (n=27)	13.12±1.55	13.73±1.67	*<0.001*
Hct			
Dapagliflozin (n=45)	40.31±4.53	41.60±4.48	*0.001*
Empagliflozin 10/25 mg (n=27)	39.77±3.99	41.94±4.50	*<0.001*
Albuminuria			
Dapagliflozin (n=45)	318.5 (0–4,817)	120.5 (0–6,259)	*0.003*
Empagliflozin 10/25 mg (n=27)	99.5 (0–3,417)	27.5 (0–4,490)	0.390
SBP			
Dapagliflozin (n=45)	150.6±21.3	137.9±18.8	*<0.001*
Empagliflozin 10/25 mg (n=27)	149.6±20.9	138.2±20.9	*0.002*
DBP			
Dapagliflozin (n=45)	80.9±12.2	75.3±10.9	*0.001*
Empagliflozin 10/25 mg (n=27)	78.4±8.0	76.7±9.2	0.333

Cre: creatinine; GFR: estimated glomerular filtration rate; Hb: haemoglobin; Hct: haematocrit; SBP: systolic blood pressure; DBP: diastolic blood pressure. Italic values indicate statistical significance (p<0.05).

SGLT2 inhibitors promote EPO secretion, which subsequently increases Hct and contributes to improved outcomes in heart failure[19,21]. Furthermore, it has been hypothesized that SGLT2 inhibitor-mediated activation of hypoxia-inducible factor 2 (HIF-2) may enhance the secretion of EPO from renal interstitial cells. This mechanism is considered a potential cause for the observed increase in Hb and Hct levels[22-24].

There is limited data in the literature on the effect of SGLT2 inhibitors on EPO levels in CKD patients. Our retrospective study was limited by the inability to collect these data. This is a notable limitation, as the decrease in EPO levels due to nephron loss is a significant cause of anaemia in CKD, particularly when the eGFR falls below 60 mL/min/1.73 m^2. Despite this, an increase in both Hb and Hct was observed in our study cohort, even among patients with an eGFR below this threshold. In a recent post hoc mediation analysis of the EMPA-REG OUTCOME trial, Wanner et al. reported that an increase in both Hb and Hct was the strongest mediator of empagliflozin's kidney benefits in patients with T2DM and CVD[25]. The EMPA-REG OUTCOME trial demonstrated that both the 10 mg and 25 mg doses of empagliflozin provided consistent risk reductions for major cardiovascular events and the progression of nephropathy. This finding suggests that the core protective mechanisms of SGLT2 inhibitors—specifically, the reduction of glomerular hyperfiltration and the mitigation of renal cortical hypoxia—are largely saturated at both dose levels. Furthermore, the existing literature indicates that the benefits of SGLT2 inhibitors in correcting anaemia can often be achieved even with lower doses, without significantly enhanced benefit from higher dosages. Consequently, this established pharmacological equivalence strongly supports our decision to combine the 10 mg and 25 mg empagliflozin cohorts into a single therapeutic category for the analysis of renal and haematological endpoints in our study.

Major trials did not report Hb and Hct data for patients with an eGFR below 60 mL/min/1.73 m^2. Our study aimed

to address this gap and observed an increase in the relevant levels even within this patient group below the eGFR threshold. This finding, coupled with Milton Packer's research on SGLT2 inhibitors activating HIF-2α, suggests that these drugs could become a future treatment option for anaemia associated with chronic DKD[26]. Further studies are needed to validate the effect of SGLT2 inhibitors on anaemia in patients with CKD, particularly in those with an eGFR below 60 mL/min/1.73 m^2. This specific subgroup warrants further investigation to confirm these findings.

The primary limitations of our investigation stem from its retrospective, single-centre design and the characteristics of the study cohort. Specifically, the absence of a dedicated control group restricts our ability to definitively attribute the observed small changes in Hb/Hct solely to SGLT2 inhibitor therapy, rather than to routine clinical management or other unmeasured factors (EPO). Crucially, our population was largely non-anaemic at baseline (mean Hb>13 g/dL) and the short duration of follow-up (3 months) limits the generalizability of our findings regarding sustained, clinically meaningful long-term effects.

In conclusion, our results indicate a potentially favourable signal on haematological parameters; however, we agree that this signal must be viewed as an area requiring further investigation rather than an established indication. Consequently, we advocate for the necessity of large-scale, high-sample, prospective, controlled studies utilizing real-world data from truly anaemic CKD populations. Such future research is essential to conclusively determine the causal role, magnitude, and clinical utility of SGLT2 inhibitor use in the management of CKD-related anaemia.

ETHICAL APPROVAL

Ethical approval and consent to participate: All procedures performed in studies involving human participants were following the ethical standards of the institutional and/or national research committee at which the studies were conducted (IRB approval number 2024/029) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all participants.

AUTHORS' CONTRIBUTIONS

SK: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Writing – original draft, Writing – review & editing. **BD:** Data curation, Formal Analysis, Funding acquisition, Writing – original draft, Writing – review & editing. **FE:** Data curation, Formal Analysis, Funding acquisition. **İG:** Data curation, Formal Analysis, Funding acquisition, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Afkarian M, Sachs MC, Kestenbaum B, Hirsch IB, Tuttle KR, Himmelfarb J, et al. Kidney disease and increased mortality risk in type 2 diabetes. J Am Soc Nephrol. 2013;24(2):302-8. https://doi.org/10.1681/ASN.2012070718
2. Hazin MAA. Anemia in chronic kidney disease. Rev Assoc Med Bras (1992). 2020;66(Suppl 1):s55-8. https://doi.org/10.1590/1806-9282.66.S1.55
3. Gohda T, Murakoshi M. Sodium-glucose cotransporter-2 inhibitors-miracle drugs for the treatment of chronic kidney disease irrespective of the diabetes status: lessons from the dedicated kidney disease-focused CREDENCE and DAPA-CKD trials. Int J Mol Sci. 2022;23(22):13749. https://doi.org/10.3390/ijms232213749
4. Maruyama T, Takashima H, Oguma H, Nakamura Y, Ohno M, Utsunomiya K, et al. Canagliflozin improves erythropoiesis in diabetes patients with anemia of chronic kidney disease. Diabetes Technol Ther. 2019;21(12):713-20. https://doi.org/10.1089/dia.2019.0212
5. Sano M, Goto S. Possible mechanism of hematocrit elevation by sodium glucose cotransporter 2 inhibitors and associated beneficial renal and cardiovascular effects. Circulation. 2019;139(17):1985-7. https://doi.org/10.1161/CIRCULATIONAHA.118.038881
6. Budzianowski J, Rzeźniczak J, Hiczkiewicz J, Kasprzak D, Winnicka-Zielińska A, Musielak B, et al. Beneficial effects of empagliflozin on hematocrit levels in a patient with severe anemia. Daru. 2021;29(2):507-10. https://doi.org/10.1007/s40199-021-00417-5
7. Koshino A, Schechter M, Chertow GM, Vart P, Jongs N, Toto RD, et al. Dapagliflozin and anemia in patients with chronic kidney disease. NEJM Evid. 2023;2(6):EVIDoa2300049. https://doi.org/10.1056/EVIDoa2300049
8. Hu JC, Shao SC, Tsai DH, Chuang AT, Liu KH, Lai EC. Use of SGLT2 inhibitors vs GLP-1 RAs and anemia in patients with diabetes and CKD. JAMA Netw Open. 2024;7(3):e240946. https://doi.org/10.1001/jamanetworkopen.2024.0946
9. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int. 2024;105(Suppl 4S):S117-314. https://doi.org/10.1016/j.kint.2023.10.018
10. Heerspink HJL, Stefánsson BV, Correa-Rotter R, Chertow GM, Greene T, Hou FF, et al. Dapagliflozin in patients with chronic

kidney disease. N Engl J Med. 2020;383(15):1436-46. https://doi.org/10.1056/NEJMoa2024816

11. Gajjala PR, Sanati M, Jankowski J. Cellular and molecular mechanisms of chronic kidney disease with diabetes mellitus and cardiovascular diseases as its comorbidities. Front Immunol. 2015;6:340. https://doi.org/10.3389/fimmu.2015.00340
12. Meda E, Pavkov AJC, Coresh J, Nelson RG. Kidney disease in diabetes in America. 3rd ed. Chapter 22. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases (US); 2018.
13. The EMPA-KIDNEY Collaborative Group, Herrington WG, Staplin N, Wanner C, Green JB, Hauske SJ, et al. Empagliflozin in patients with chronic kidney disease. N Engl J Med. 2022;388(2):117-27. https://doi.org/10.1056/NEJMoa2204233
14. Podestà MA, Sabiu G, Galassi A, Ciceri P, Cozzolino M. SGLT2 inhibitors in diabetic and non-diabetic chronic kidney disease. Biomedicines. 2023;11(2):279. https://doi.org/10.3390/biomedicines11020279
15. Griffin M, Rao VS, Ivey-Miranda J, Fleming J, Mahoney D, Maulion C, et al. Empagliflozin in heart failure: diuretic and cardiorenal effects. Circulation. 2020;142(11):1028-39. https://doi.org/10.1161/CIRCULATIONAHA.120.045691
16. Jensen J, Omar M, Kistorp C, Poulsen MK, Tuxen C, Gustafsson I, et al. Twelve weeks of treatment with empagliflozin in patients with heart failure and reduced ejection fraction: a double-blinded, randomized, and placebo-controlled trial. Am Heart J. 2020;228:228:47-56. https://doi.org/10.1016/j.ahj.2020.07.011
17. McMurray JJV, Solomon SD, Inzucchi SE, Køber L, Kosiborod MN, Martinez FA, et al. Dapagliflozin in patients with heart failure and reduced ejection fraction. N Engl J Med. 2019;381(21):1995-2008. https://doi.org/10.1056/NEJMoa1911303
18. Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, et al. Cardiovascular and renal outcomes with empagliflozin in heart failure. N Engl J Med. 2020;383(15):1413-24. https://doi.org/10.1056/NEJMoa2022190
19. Vallon V, Verma S. Effects of SGLT2 inhibitors on kidney and cardiovascular function. Annu Rev Physiol. 2021;83:503-28. https://doi.org/10.1146/annurev-physiol-031620-095920
20. Perkovic V, Jardine MJ, Neal B, Bompoint S, Heerspink HJL, Charytan DM, et al. Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. N Engl J Med. 2019;380(24):2295-306. https://doi.org/10.1056/NEJMoa1811744
21. Lambers Heerspink HJ, Zeeuw D, Wie L, Leslie B, List J. Dapagliflozin a glucose-regulating drug with diuretic properties in subjects with type 2 diabetes. Diabetes Obes Metab. 2013;15(9):853-62. https://doi.org/10.1111/dom.12127
22. Onishi A, Fu Y, Patel R, Darshi M, Crespo-Masip M, Huang W, et al. A role for tubular Na+/H+ exchanger NHE3 in the natriuretic effect of the SGLT2 inhibitor empagliflozin. Am J Physiol Renal Physiol. 2020;319(4):F712-28. https://doi.org/10.1152/ajprenal.00264.2020
23. Mazer CD, Hare GMT, Connelly PW, Gilbert RE, Shehata N, Quan A, et al. Effect of empagliflozin on erythropoietin levels, iron stores, and red blood cell morphology in patients with type 2 diabetes mellitus and coronary artery disease. Circulation. 2020;141(8):704-7. https://doi.org/10.1161/CIRCULATIONAHA.119.044235
24. Inzucchi SE, Zinman B, Fitchett D, Wanner C, Ferrannini E, Schumacher M, et al. How does empagliflozin reduce cardiovascular mortality? Insights from a mediation analysis of the EMPA-REG OUTCOME trial. Diabetes Care. 2018;41(2):356-63. https://doi.org/10.2337/dc17-1096
25. Wanner C, Nangaku M, Kraus BJ, Zinman B, Mattheus M, Hantel S, et al. How do SGLT2 inhibitors protect the kidney? A mediation analysis of the EMPA-REG OUTCOME trial. Nephrol Dial Transplant. 2024;39(9):1504-13. https://doi.org/10.1093/ndt/gfae032
26. Packer M. Mechanistic and clinical comparison of the erythropoietic effects of SGLT2 inhibitors and prolyl hydroxylase inhibitors in patients with chronic kidney disease and renal anemia. Am J Nephrol. 2024;55(2):255-9. https://doi.org/10.1159/000531084

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251462

Whole-body magnetic resonance imaging vs. positron emission tomography-computed tomography in pediatric oncology: enhancing diagnostic precision for solid tumors

Süreyya Burcu Görkem[1,2]*, Gökalp Çıkman[1,3]

SUMMARY

OBJECTIVE: Pediatric solid tumors require accurate imaging for staging, treatment planning, and follow-up. Whole-body magnetic resonance imaging has emerged as a promising radiation-free alternative to positron emission tomography-computed tomography, offering superior soft tissue contrast. This study compares the diagnostic performance of whole-body magnetic resonance imaging and positron emission tomography-computed tomography in detecting primary tumors, metastases, and recurrences in pediatric solid tumors.

METHODS: This retrospective study included 47 pediatric patients (27 boys and 20 girls; age range: 3–18 years; mean age: 11.3±3.5 years) with suspected or confirmed solid tumors who underwent both whole-body magnetic resonance imaging and positron emission tomography-computed tomography between 2017 and 2020. Lesional and per-patient agreement between modalities was evaluated using positron emission tomography-computed tomography as the reference. Statistical analysis was performed to assess inter-modality agreement.

RESULTS: Positron emission tomography-computed tomography detected pathological findings in 35 patients (74.4%), including 27 primary tumors (57.4%), 18 metastases (38.2%), and one recurrence (2.1%). Whole-body magnetic resonance imaging showed high diagnostic concordance, detecting all primary tumors and the recurrence, and 16 of 18 metastatic cases. Per-lesion analysis revealed 66 lesions on positron emission tomography-computed tomography and 64 on whole-body magnetic resonance imaging, with two metastatic lesions (an ossified pulmonary metastasis and an osteoblastic femur lesion) missed by magnetic resonance imaging. Cohen's kappa (κ) for abnormality detection was 0.89, indicating almost perfect agreement (95%CI 0.75–1.00). McNemar's test showed no statistically significant difference between the modalities (p=0.479).

CONCLUSION: Whole-body magnetic resonance imaging demonstrates excellent agreement with positron emission tomography-computed tomography for evaluating pediatric solid tumors, offering a radiation-free alternative, particularly advantageous in long-term follow-up. While limitations remain in detecting sclerotic metastases, whole-body magnetic resonance imaging is a primary alternative in specific scenarios.

KEYWORDS: PET-CT. Diagnostic imaging. Comparative study. Child. Adolescent. Neoplasms. Whole body imaging.

INTRODUCTION

Pediatric solid tumors are a heterogeneous group of malignancies that arise from various tissues and organs, including bone, soft tissue, and organs such as the liver, kidney, and brain. These tumors are distinct from adult cancers in terms of their histology, biology, and clinical behavior, necessitating a specialized approach to their diagnosis and management[1]. Accurate assessment of these tumors is critical for determining the appropriate therapeutic strategies and improving patient outcomes. Given the aggressive nature of many pediatric cancers and the risk of metastasis, comprehensive imaging plays a pivotal role in staging disease treatment, treatment planning, response monitoring, and long-term follow-up. Positron emission tomography combined with computed tomography (PET-CT) has been the cornerstone imaging modality for evaluating pediatric solid tumors. PET-CT provides valuable metabolic and anatomical information, facilitating the detection of primary tumors, metastatic spread, and treatment responses[2]. However, the cumulative radiation dose from repeated PET-CT scans during cancer management can further heighten the risk of secondary malignancies, emphasizing the need for safer imaging alternatives. WB-MRI offers superior soft tissue contrast, enabling detailed visualization of primary tumors, bone marrow metastases, and lymph node involvement without the use of ionizing radiation. It offers a valuable radiation-free imaging option in pediatric oncology, where minimizing cumulative radiation is critical. It is particularly effective in detecting and monitoring malignancies such as neuroblastoma, Ewing sarcoma,

[1]Erciyes University, School of Medicine, Department of Pediatric Radiology – Kayseri, Turkey.
[2]Adana City Training and Research Hospital, Pediatric Radiology Clinic – Adana, Turkey.
[3]University of Toronto, The Hospital for Sick Children, Department of Pediatric Neuroradiology – Toronto (ON), Canada.
*Corresponding author: drburcugorkem@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on January 11, 2026. Accepted on February 21, 2026.
Scientific Editor: José Maria Soares Júnior

rhabdomyosarcoma, and lymphoma, supporting comprehensive assessment for staging, metastasis screening, and follow-up. However, limitations include longer acquisition times that may require sedation in younger patients, as well as higher costs and limited availability. Recent advancements in MRI technology, including faster sequences and improved whole-body imaging protocols, continue to enhance its practicality and integration into clinical practice[3]. The clinical significance of comparing WB-MRI to PET-CT lies in determining whether WB-MRI can match or surpass the diagnostic performance of PET-CT while mitigating radiation risks. By highlighting the advantages and limitations of both imaging techniques, this study aims to provide valuable insights into the evolving role of WB-MRI in pediatric oncology and its potential to become a standard imaging modality in the comprehensive care of children with solid tumors.

METHODS

The institutional ethical board (Erciyes University School of Medicine, 2020/367; 07/08/2020) approved this retrospective study.

Study design and patient selection

This retrospective study included 47 pediatric patients (27 boys and 20 girls; age range: 3–18 years; mean age: 11.3±3.5 years) with suspected or confirmed solid tumors who underwent both WB-MRI and PET-CT between 2017 and 2020 (Table 1). Patients were identified through a retrospective review of institutional imaging and oncology records. The decision to perform both WB-MRI and PET-CT was not based on a standardized institutional protocol but was determined at the discretion of the managing oncology team, depending on clinical factors such as diagnostic uncertainty, the need for comprehensive staging in high-risk or atypical tumors, or complex follow-up scenarios where additional imaging clarity was required. This non-random selection may introduce selection bias and is acknowledged as a limitation of the study. Inclusion criteria were as follows: (a) patients with histologically confirmed or clinically suspected primary solid tumors, (b) patients with known or suspected metastases or recurrence at the time of evaluation, (c) patients undergoing initial staging or follow-up evaluation who had both PET-CT and WB-MRI performed within a clinically relevant time window (≤2 weeks apart), (d) patients with no evidence of residual disease on follow-up imaging or post-surgical pathology. Our exclusion criterion was patients who underwent chemotherapy, radiotherapy, or surgical intervention after PET-CT but before WB-MRI (n=5), as such treatments could alter imaging findings and confound comparative assessment.

Positron emission tomography-computed tomography protocol

PET-CT was performed by the Philips Gemini TF 64 PET-CT device [tumor imaging procedure according to the European Society for Nuclear Medicine (EANM) guideline version 2.0.4][4]. Patients' oral intake was stopped at least 4 h before the 18F-FDG injection. In addition, blood glucose was checked before the extraction. In PET-CT, 5–7 MBq/kg of fluorodeoxyglucose was used intravenously as a radiopharmaceutical. Images were taken from the vertex of the head to the heels 60–80 min after the 18F-FDG injection using the 3D application. The total examination time in the PET-CT room was between 1.5 and 2 h.

Whole-body magnetic resonance imaging protocol

All MR images were performed by a 1.5 T MRI unit (Siemens, MAGNETOM Aera, Germany). For each patient, from head to toe, coronal and axial or sagittal images were obtained in short tau inversion recovery (STIR) sequence [TR (ms)/TE (ms)=5,000/75; inversion time=160 ms to null fat sign; slice thickness=4–5 mm; slice gap value 1 mm; matrix value= 256×256, and field of view (FOV) coronal=350–400 mm and 200–400 mm, for axial images] without administration of any contrast agent by using standard head and neck coils, standard posterior spine coil, and 2× anterior body coils for thorax, abdomen, and limbs. Patients were told to use the toilet before the examination. Younger children who could not follow the instructions received sedation

Table 1. Distribution of pediatric solid tumors in the study cohort (n=47).

Tumor	Number of patients
Ewing sarcoma	10
Neuroblastoma	9
Osteosarcoma	6
PNET	4
Rhabdomyosarcoma	4
Adrenocortical tumor	4
Synovial sarcoma	2
Ganglioneuroblastoma	1
Germ cell tumor	1
Hepatoblastoma	1
Immature teratoma	1
Infantile fibrosarcoma	1
Colon adenocarcinoma	1
Malignant melanoma	1
Askin's tumor	1

PNET, primitive neuroectodermal tumor.

(intranasal 0.1–0.2 mg/kg, Dormicum, Roche). The patient's position was head-first with arms down by their side or crossed on their body. The total examination time in the MRI room was between 20 and 45 min.

Image evaluation

Two pediatric radiologists (SBG and GC) jointly reviewed all WB-MRI examinations in a consensus reading session. The reviewers were blinded to patient clinical data and PET-CT findings to minimize interpretation bias. WB-MRI images were evaluated using a dedicated PACS workstation (Picture Archiving and Communication Systems, Sectra IDS7), and findings were subsequently compared with the corresponding PET-CT reports.

Statistical analysis

All statistical analyses were performed using IBM SPSS Statistics version 22 (IBM SPSS Inc., Chicago, IL, USA). The distribution of continuous variables was assessed using the Kolmogorov-Smirnov test. To compare the diagnostic performance of the two imaging modalities, McNemar's test was employed for analyzing paired binary outcomes, such as differences in detection rates, reflecting whether the proportions of positive findings differed significantly between modalities. To evaluate the agreement between the two imaging modalities beyond chance, Cohen's kappa coefficient was calculated, providing a measure of inter-modality reliability. A p-value less than 0.05 was considered statistically significant for all tests.

RESULTS

In total, 47 children diagnosed with solid tumors were assessed through a retrospective review. Among these, 35 patients (74.4%) demonstrated pathological findings on PET-CT, including 27 with primary tumors (57.4%), 18 with metastases (38.2%), and one case with postoperative recurrence (2.1%). The remaining 12 patients (25.5%) exhibited no pathological findings on PET-CT and WB-MRI at either initial diagnosis or during follow-up, indicating complete concordance. WB-MRI showed high diagnostic concordance, detecting all primary tumors and the recurrence, and 16 of 18 metastatic cases. Notably, 28 patients underwent whole-body MRI (WB-MRI) at the time of initial diagnosis and 19 patients received both initial and follow-up WB-MRI studies. Between two and nine WB-MRI examinations were performed per patient, with a maximum follow-up duration of 1 year. The overall kappa value (κ) for any abnormality detection was 0.89 (95%CI 0.75–1.00), which indicates almost perfect agreement (95.7%) between WB-MRI and PET-CT per patient. McNemar's yielded a p-value of 0.479, indicating no statistically significant difference in malignancy detection between the two imaging modalities per patient.

On a per-lesion basis, a total of 66 lesions were detected on PET-CT, comprising 27 primary tumors (40.9%), 10 nodal metastases (15%) located in cervical, axillary, mediastinal, hilar, and abdominal regions; 12 bone or bone marrow metastases (18%); 5 peritoneal or retroperitoneal metastases (7.5%); 11 visceral metastases [7 pulmonary, 3 hepatic, and 1 central nervous system (16.6%)]; and 1 postoperative recurrence (1.5%). WB-MRI identified a nearly identical lesion distribution with a total of 64 lesions, including 27 primary tumors (42.1%), 10 nodal metastases (15.6%), 11 bone or bone marrow metastases (17%), 5 peritoneal/retroperitoneal metastases (7.8%), and 10 visceral metastases [6 lung, 3 liver, 1 central nervous system (15.6%)], as well as the single recurrence case (1.5%). WB-MRI failed to identify two lesions seen on PET-CT: an osteoblastic femur metastasis and an ossified lung metastasis. In each case, it resulted in a change in clinical staging from localized to metastatic disease, leading to escalation of chemotherapy intensity and inclusion of curative protocols such as localized surgical resection or focal radiotherapy. A comprehensive summary of patient-level imaging agreement is presented in Table 2. Comparative examples illustrating modality-specific findings and representative cases are shown in Figures 1A–D, 2A–C, 3A, B, and 4A–D.

DISCUSSION

This study highlights the potential of WB-MRI as a valuable imaging modality in the evaluation of pediatric solid tumors. By retrospectively comparing WB-MRI findings to those of PET-CT—a widely used reference in oncologic images, we assessed the diagnostic concordance between the two modalities in detecting primary tumors, metastases, and recurrences. Our findings demonstrated a high degree of agreement between WB-MRI and PET-CT (κ=0.89, 95%CI 0.75–1.00) when PET-CT was used as the comparator. These figures indicate that WB-MRI often matches other modalities, but they show agreement between methods, not accuracy. This near-equivalent performance is particularly noteworthy given the absence of ionizing radiation in WB-MRI, which is a significant consideration in pediatric populations vulnerable to cumulative radiation risks. Given that many pediatric oncology protocols require repeated imaging, WB-MRI offers a compelling alternative in longitudinal care strategies[5]. Moreover, the ability of WB-MRI to detect the majority of metastatic lesions underscores its robustness and supports further evaluation in prospective, protocol-optimized settings.

Table 2. Lesion-based comparison of positron emission tomography-computed tomography and whole-body magnetic resonance imaging findings in pediatric solid tumors.

Imaging findings	PET-CT (n, %)	WB-MRI (n, %)	Agreement (%)	p-value (McNemar's)
Primary tumor	27 (40.9%)	27 (42.6%)	100%	N/A
Lymphadenopathy	10 (15%)	10 (15.6%)	100%	N/A
Bone/bone marrow metastasis	12 (18%)	11 (17%)	91.6%	*p>0.05*
Peritoneal/retroperitoneal metastasis	5 (7.5%)	5 (7.8%)	100%	N/A
Visceral metastasis	11 (16.6%)	10 (15.6%)	90.9%	*p>0.05*
Postoperative recurrence	1 (1.5%)	1 (1.5%)	100%	N/A
All findings (total)	66 (100%)	64 (100%)	96.9%	0.479

Note: Agreement values are based on per-lesion analysis. Percentages reflect the proportion of each lesion type relative to total PET-CT and WB-MRI detections. McNemar's test p-values>0.05 indicate no statistically significant difference in detection rates between modalities. PET-CT: positron emission tomography-computed tomography; WB-MRI: whole-body magnetic resonance imaging.

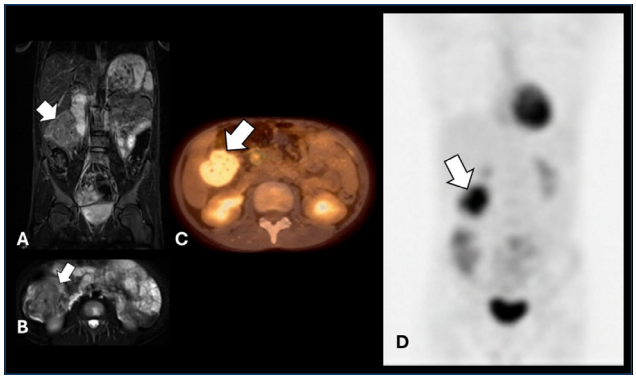


Figure 1. (A–D) 16-year-old boy with colon adenocarcinoma: coronal **(A)** and axial short tau inversion recovery images **(B)** of whole-body magnetic resonance imaging through the abdomen demonstrate heterogeneous hyperintense mass in the right hemicolon (white arrows). Positron emission tomography-computed tomography fusion axial image **(C)** indicates increased Fludeoxyglucose F 18 uptake (white arrow) in the mass. Positron emission tomography-computed tomography metabolic map image **(D)** shows Fludeoxyglucose F 18 uptake compared to surrounding tissues.

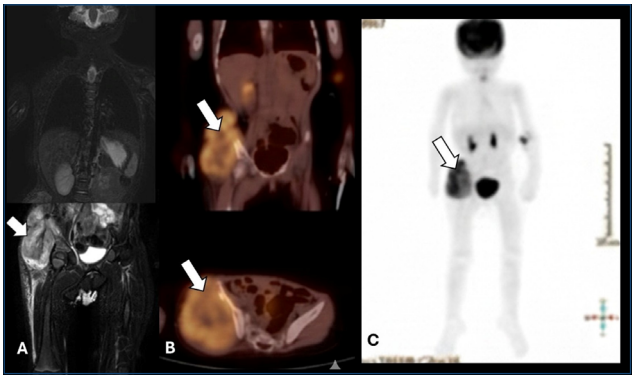


Figure 2. (A–C) 3-year-old girl infantile fibrosarcoma: coronal whole-body magnetic resonance imaging short tau inversion recovery image **(A)** demonstrates the right hemipelvic hyperintense solid mass (white arrow). Positron emission tomography-computed tomography fusion coronal and axial images show the right hemipelvic solid mass (white arrows) with an increased Fludeoxyglucose F 18 uptake **(B)**. Positron emission tomography-computed tomography metabolic map image **(C)** shows Fludeoxyglucose F 18 uptake compared to surrounding tissues (white arrow).

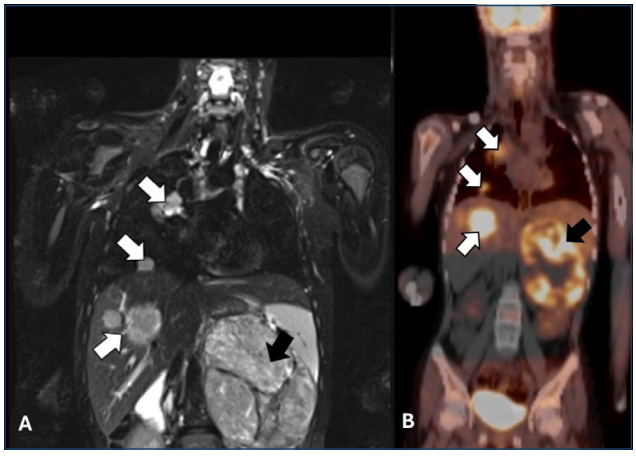


Figure 3. (A,B) 18-year-old girl with left adrenocortical carcinoma: Coronal whole-body magnetic resonance imaging short tau inversion recovery image **(A)** shows the hyperintense solid mass in the left adrenal region (black arrow) associated with multiple liver and lung metastases and hilar lymphadenopathies (white arrows). positron emission tomography-computed tomography image **(B)** demonstrates the same findings with increased Fludeoxyglucose F 18 uptakes (white arrows and black arrow).

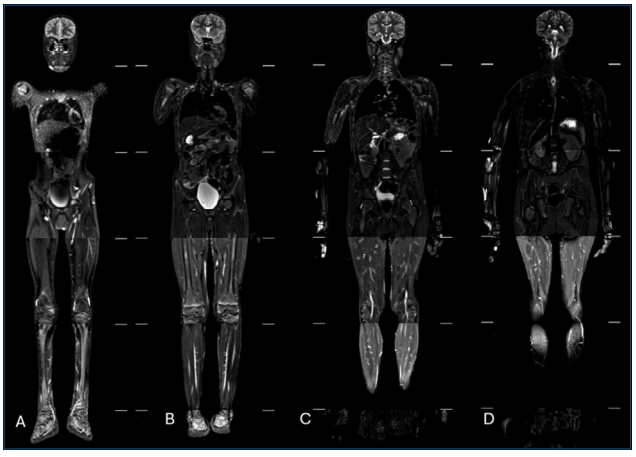


Figure 4. (A–D) 12-year-old boy with synovial sarcoma: consecutive coronal whole-body magnetic resonance imaging short tau inversion recovery images **(A–D)** through whole body show no metastasis or recurrence on follow-up.

Currently, standard imaging procedures for neuroblastomas and sarcomas involve a combination of CT, MRI, I-123 MIBG scintigraphy, and PET-CT, depending on the patient's specific needs[6]. Our findings suggest that WB-MRI could potentially replace or complement these modalities in selected cases, especially when reducing radiation burden is a clinical priority. Nevertheless, a multidisciplinary imaging approach remains essential, and WB-MRI should be considered as part of a broader diagnostic framework rather than a standalone replacement[7].

WB-MRI also plays a valuable role in monitoring tumor response during neoadjuvant chemotherapy or radiotherapy, offering precise anatomical detail that can support surgical planning and guide ongoing treatment decisions. Due to its high soft tissue resolution and lack of ionizing radiation, WB-MRI is particularly well-suited for repeated use in pediatric patients. Although PET-CT remains essential for assessing metabolic activity and identifying certain metastatic patterns, its effectiveness in evaluating treatment response is limited to changes in metabolic uptake, which may not always correlate with structural tumor changes. In contrast, WB-MRI enables direct visualization of tumor size, extent, and tissue characteristics, making it a strong candidate for incorporation into both diagnostic and follow-up protocols, especially when minimizing radiation exposure is a priority. Rather than serving as a replacement, WB-MRI should be viewed as a complementary tool to PET-CT, with advantages in preoperative assessment, treatment monitoring, and long-term surveillance in pediatric oncology patients[8,9]. WB-MRI demonstrates considerable clinical value in the assessment of specific pediatric tumor types, including neuroblastoma, Ewing sarcoma, rhabdomyosarcoma, and lymphoma. It allows for comprehensive anatomical evaluation, including detection and characterization of primary tumors, assessment of bone marrow involvement, and monitoring of treatment response over time. The STIR sequence is central to WB-MRI, providing effective fat suppression and enhancing lesion visibility in bone marrow and soft tissues[10-12]. The two metastatic lesions missed by WB-MRI—an osteoblastic femoral metastasis and an ossified pulmonary nodule—reflect inherent limitations in MRI physics and common whole-body sequences such as STIR. These sequences are sensitive to lesions with high water content but are less effective in detecting dense, calcified, or sclerotic tissue. Osteoblastic metastases have low proton density and short T2 relaxation times due to their high mineral content, resulting in minimal signal on both STIR and conventional MR images. Similarly, calcified lung nodules contain little mobile hydrogen and often appear hypointense or are completely inapparent on MRI. In STIR sequences, fat suppression can further obscure such lesions, particularly when they are small or embedded within low-signal structures like bone or lung parenchyma. These technical constraints explain the false-negative findings in our study and highlight the need for complementary imaging approaches when such metastatic patterns are suspected. PET-CT provides superior sensitivity in detecting certain types of metastases, most notably small pulmonary nodules, which may be difficult to visualize on MRI due to motion artifacts and limited lung parenchyma contrast. Additionally, PET-CT enables quantitative metabolic assessment through standardized uptake values (SUV), which can be instrumental in evaluating treatment response, monitoring minimal residual disease, and guiding biopsy targets. PET-CT examinations also typically have shorter acquisition times and often do not require sedation, making them more practical in certain clinical settings, particularly for younger or uncooperative children[4]. In contrast, WB-MRI's strength lies in its high soft tissue resolution, spinal and bone marrow evaluation, and radiation-free follow-up potential—factors that are particularly valuable in pediatric patients requiring serial imaging. Given these complementary advantages, the two modalities may be best viewed as tools with overlapping and distinct roles. Future clinical algorithms could benefit from individualized modality selection based on tumor type, disease distribution, patient age, and specific diagnostic goals.

Currently, no universal imaging protocol exists for pediatric solid tumors, but the addition of sequences such as diffusion-weighted imaging (DWI) and T1-weighted imaging enhances WB-MRI's diagnostic performance, especially for identifying bone marrow and lymph node abnormalities[13,14]. Advanced techniques like perfusion imaging can further support evaluation of tissue vascularity and viability. Nevertheless, challenges remain—particularly the need for longer acquisition times and potential sedation in younger children, which can limit its practicality in some settings.

Recent studies suggest that streamlined WB-MRI protocols using rapid 3D STIR sequences may reduce scan time, improve patient compliance, and minimize the need for sedation without sacrificing diagnostic quality[10]. While our findings and existing literature support WB-MRI's utility, especially in radiation-sensitive pediatric populations, its reduced sensitivity for detecting certain calcified or ossified metastases remains a limitation. This highlights the importance of continued refinement of WB-MRI protocols and reinforces the role of PET-CT as a complementary modality in providing a more complete diagnostic assessment.

Future directions and proposed whole-body magnetic resonance imaging protocol

Given the high concordance observed in this study, future research should aim to refine and standardize WB-MRI protocols to enhance diagnostic accuracy and clinical applicability. One key area of improvement lies in the integration of advanced sequences that

are now widely used in modern WB-MRI practices. A proposed optimized protocol would include (1) coronal and sagittal STIR sequences for broad soft tissue and bone marrow screening; (2) axial DWI with background suppression, such as DWIBS (diffusion-weighted whole-body imaging with background body signal suppression), to improve sensitivity for small or early lesions; and (3) axial T1-weighted sequences for detailed anatomical localization and marrow characterization. The addition of contrast-enhanced sequences may be considered in selected cases where vascularity or lesion perfusion is critical, although contrast-free protocols retain advantages in safety, efficiency, and repeatability for longitudinal follow-up. Standardizing such a protocol across institutions would improve comparability, reduce variability, and better position WB-MRI as a primary imaging modality in pediatric oncology. Future prospective studies using this comprehensive protocol, ideally with histopathologic or long-term follow-up confirmation, are essential to validate and expand upon the findings of this retrospective analysis.

Limitations

Despite its strengths, this study has certain limitations. The retrospective design introduces potential selection bias, as the choice to perform both WB-MRI and PET-CT was based on clinical discretion rather than a standardized protocol. Additionally, the absence of an independent reference standard, such as histopathology or long-term clinical follow-up, limits the ability to calculate true diagnostic accuracy. The lack of such a reference standard limits our ability to make definitive claims regarding the absolute diagnostic accuracy of either modality. The patient population was also heterogeneous in terms of tumor types and imaging timing (initial vs. follow-up), which may affect the generalizability of the findings. From a technical standpoint, WB-MRI failed to detect two specific ossified metastases due to limitations of standard MR sequences such as STIR, which is less sensitive to ossified low-proton-density lesions. Furthermore, we used only STIR sequences without DWI or contrast-enhanced sequences, which limits the generalizability of the results, as modern WB-MRI protocols often incorporate DWI to enhance sensitivity for small or subtle lesions and gadolinium contrast to improve lesion conspicuity and tissue characterization. Practical challenges also remain, including long scan times, the need for sedation in younger children, and limited availability in some healthcare settings.

CONCLUSION

WB-MRI demonstrated high concordance and comparable diagnostic performance with PET-CT, avoiding radiation exposure in the evaluation of primary tumors, metastatic disease, and recurrence in pediatric solid tumors. Its ability to provide comprehensive, radiation-free whole-body imaging makes it a valuable modality in pediatric oncology, particularly where long-term surveillance is required. Although limitations exist—particularly in detecting sclerotic or calcified lesions—WB-MRI remains effective when appropriate sequences such as STIR and DWI are utilized.

Within the context of modern pediatric oncology practice, WB-MRI offers a complementary approach to existing imaging strategies. It enhances soft tissue assessment and supports safer longitudinal imaging by reducing cumulative radiation exposure. With ongoing technological advancements and expanding availability, WB-MRI is poised to become an increasingly vital component of staging, treatment planning, and longitudinal follow-up in pediatric patients with solid tumors. Prospective studies are essential to further validate its utility and refine its clinical indications.

ACKNOWLEDGMENTS

This study was accepted as a poster presentation: "Comparison of Whole-Body MRI and PET-CT in Pediatric Solid Tumors" at the International Pediatric Radiology Congress (IPR) on October 11–15, 2021, held in Rome, ITALY.

DECLARATION OF AI-ASSISTED TECHNOLOGIES

We confirm that all scientific data and results in this study were generated and analyzed without the use of AI tools. OpenAI's ChatGPT contributed to refining the manuscript by helping with language editing, organization, and enhancing its overall clarity. The authors created, reviewed, and approved all the scientific material, and they fully accept responsibility for the work's accuracy and integrity.

AUTHORS' CONTRIBUTIONS

SBG: Conceptualization, Data curation, Formal Analysis, Writing – original draft. **GÇ:** Data curation, Formal Analysis.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Goo HW. Whole-body MRI in children: current imaging techniques and clinical applications. Korean J Radiol. 2015;16(5):973-85. https://doi.org/10.3348/kjr.2015.16.5.973
2. Schäfer JF, Granata C, Kalle T, Kyncl M, Littooij AS, Paolo PL, et al. Whole-body magnetic resonance imaging in pediatric oncology - recommendations by the Oncology Task Force of the ESPR. Pediatr Radiol. 2020;50(8):1162-74. https://doi.org/10.1007/s00247-020-04683-4
3. Zadig P, Brandis E, Lein RK, Rosendahl K, Avenarius D, Ording Müller LS. Whole-body magnetic resonance imaging in children - how and why? A systematic review. Pediatr Radiol. 2021;51(1):14-24. https://doi.org/10.1007/s00247-020-04735-9
4. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42(2):328-54. https://doi.org/10.1007/s00259-014-2961-x
5. Gottumukkala RV, Gee MS, Hampilos PJ, Greer MC. Current and emerging roles of whole-body MRI in evaluation of pediatric cancer patients. Radiographics. 2019;39(2):516-34. https://doi.org/10.1148/rg.2019180130
6. Greer MC, Voss SD, States LJ. Pediatric cancer predisposition imaging: focus on whole-body MRI. Clin Cancer Res. 2017;23(11):e6-13. https://doi.org/10.1158/1078-0432.CCR-17-0515
7. Davis JT, Kwatra N, Schooler GR. Pediatric whole-body MRI: a review of current imaging techniques and clinical applications. J Magn Reson Imaging. 2016;44(4):783-93. https://doi.org/10.1002/jmri.25259
8. Eutsler EP, Khanna G. Whole-body magnetic resonance imaging in children: technique and clinical applications. Pediatr Radiol. 2016;46(6):858-72. https://doi.org/10.1007/s00247-016-3586-y
9. Nievelstein RA, Littooij AS. Whole-body MRI in paediatric oncology. Radiol Med. 2016;121(5):442-53. https://doi.org/10.1007/s11547-015-0600-7
10. Merlini L, Carpentier M, Ferrey S, Anooshiravani M, Poletti PA, Hanquinet S. Whole-body MRI in children: would a 3D STIR sequence alone be sufficient for investigating common paediatric conditions? A comparative study. Eur J Radiol. 2017;88:155-62. https://doi.org/10.1016/j.ejrad.2017.01.014
11. Schooler GR, Davis JT, Daldrup-Link HE, Frush DP. Current utilization and procedural practices in pediatric whole-body MRI. Pediatr Radiol. 2018;48(8):1101-7. https://doi.org/10.1007/s00247-018-4145-5
12. Anupindi SA, Bedoya MA, Lindell RB, Rambhatla SJ, Zelley K, Nichols KE, et al. Diagnostic performance of whole-body MRI as a tool for cancer screening in children with genetic cancer-predisposing conditions. AJR Am J Roentgenol. 2015;205(2):400-8. https://doi.org/10.2214/AJR.14.13663
13. Siegel MJ, Acharyya S, Hoffer FA, Wyly JB, Friedmann AM, Snyder BS, et al. Whole-body MR imaging for staging of malignant tumors in pediatric patients: results of the American College of Radiology Imaging Network 6660 trial. Radiology. 2013;266(2):599-609. https://doi.org/10.1148/radiol.12112531
14. Chavhan GB, Babyn PS. Whole-body MR imaging in children: principles, technique, current applications, and future directions. Radiographics. 2011;31(6):1757-72. https://doi.org/10.1148/rg.316115523

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251467

Relationship between lower urinary system symptoms and urinary incontinence in women with gestational diabetes

Leyla Kaya[1]*, Esra Keles[2,3], Zahide Kaya[4], Hakan Dokumuș[5]

SUMMARY

OBJECTIVE: The aim of this study was to evaluate the association between gestational diabetes mellitus and the severity and impact of lower urinary tract symptoms and urinary incontinence on quality of life in pregnant women.

METHODS: This case–control study enrolled 500 pregnant women (251 with gestational diabetes mellitus; 249 healthy controls) between 30 and 38 weeks' gestation. Women completed the International Consultation on Incontinence Questionnaire-Short Form, Urogenital Distress Inventory-Short Form, Incontinence Impact Questionnaire-Short Form, and Incontinence Severity Index. Sociodemographic, obstetrics, and anthropometric data were also collected. Multivariable logistic regression analyses were performed to examine the independent association between gestational diabetes mellitus and lower urinary tract symptoms and urinary incontinence.

RESULTS: The gestational diabetes mellitus group exhibited a significantly higher mean age compared to the control group ($p<0.001$). Pre-pregnancy body mass index and body mass index during pregnancy were also significantly elevated in the gestational diabetes mellitus group ($p<0.001$). The median gestational weight gain was significantly greater in the gestational diabetes mellitus group ($p=0.027$). Women with gestational diabetes mellitus demonstrated significantly elevated scores in International Consultation on Incontinence Questionnaire-Short Form, Urogenital Distress Inventory-6, Incontinence Impact Questionnaire-7, and Incontinence Severity Index ($p<0.001$ for all). After adjusting for age, pre-pregnancy body mass index, body mass index during pregnancy, and gestational weight gain in multivariable logistic regression models, gestational diabetes mellitus remained independently associated with increased UI risk (adjusted OR 2.34, 95%CI 1.56–3.51, $p<0.001$).

CONCLUSION: Gestational diabetes mellitus is significantly associated with increased severity of lower urinary tract symptoms and urinary incontinence, as well as a greater negative impact on quality of life during pregnancy. Routine systematic screening for lower urinary tract symptoms and urinary incontinence may be considered for all pregnant women with gestational diabetes mellitus, with early implementation of preventive interventions including supervised pelvic floor muscle training programs.

KEYWORDS: Gestational diabetes mellitus. Urinary incontinence. Quality of life. Pregnancy. Lower urinary tract symptoms. Pelvic floor disorders.

INTRODUCTION

Gestational diabetes mellitus (GDM), characterized by the onset or initial diagnosis of glucose dysregulation during pregnancy, represents the most prevalent metabolic complication impacting approximately 1–14% of pregnancies globally, with prevalence rates increasing in parallel with rising trends in obesity and advanced maternal age[1]. GDM confers significant short-term maternal and fetal risks, including macrosomia, preeclampsia, polyhydramnios, and neonatal hypoglycemia. Furthermore, it elevates long-term maternal susceptibility to Type 2 diabetes mellitus (T2DM) and cardiovascular disease[2].

Concurrently, lower urinary tract symptoms (LUTS), encompassing a range of bothersome conditions including urinary incontinence (UI), urgency, frequency, and nocturia, are exceedingly common during pregnancy[3]. The physiological adaptations of pregnancy, such as increased intra-abdominal pressure exerted by the gravid uterus on the bladder and pelvic floor, hormonal changes affecting connective tissue and muscle tone, and potential neurogenic alterations, contribute significantly to the development or exacerbation of these symptoms[4].

Existing literature has established a clear association between T2DM and heightened UI risk in females[5]. The proposed mechanisms include diabetic neuropathy affecting bladder innervation and sphincter function, diabetic myopathy impacting pelvic floor muscle (PFM) strength, and microvascular complications affecting pelvic tissue integrity[6]. Given the metabolic similarities and shared risk factors (such as obesity and advanced age) between GDM and T2DM, it is plausible that GDM may also

[1]University of Health Sciences, Faculty of Health Sciences, Department of Midwifery – Istanbul, Turkey.
[2]University of Health Sciences, Kartal Lütfi Kırdar City Hospital, Department of Gynecologic Oncology – Istanbul, Turkey.
[3]University of Health Sciences, Hamidiye Faculty of Medicine, Department of Public Health – Istanbul, Turkey.
[4]Uskudar State Hospital, Department of Internal Medicine – Istanbul, Turkey.
[5]Mersin University, Institute of Health Sciences, Department of Nursing – Mersin, Turkey.
*Corresponding author: leylakaya02@hotmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on August 01, 2025. Accepted on January 07, 2026.
Scientific Editor: José Maria Soares Júnior

predispose pregnant women to LUTS/UI or exacerbate existing symptoms[7]. However, the direct link between GDM and UI during the index pregnancy is less definitively established compared to the long-term postpartum period[8]. Reported prevalence rates of LUTS/UI in pregnant women with GDM range from 35 to 68%, depending on diagnostic criteria, population characteristics, and assessment instruments employed[9-11]. Cross-cultural studies suggest that while the GDM–LUTS/UI association appears consistent across populations, the magnitude of effect may be influenced by population-specific factors including ethnicity, healthcare access, and cultural attitudes toward symptom reporting.

Given the significant prevalence of both GDM and LUTS in pregnancy, and the potential impact on maternal well-being, a comprehensive understanding of the association between these two conditions is crucial. This study aimed to investigate the severity and impact on quality of life (QoL) of LUTS and UI in pregnant women with GDM compared to those with uncomplicated pregnancies, utilizing a battery of validated questionnaires to provide a quantitative assessment.

METHODS

This case–control study was conducted at Zeynep Kamil Women's and Children's Disease Research and Training Hospital, between April 2023 and April 2024. Pregnant women diagnosed with GDM (case group) were prospectively enrolled and compared to a healthy gestational age-matched control group receiving routine antenatal care during the same period.

Inclusion criteria for both groups included singleton pregnancy, ability to understand and complete the study questionnaires, and gestational age between 30 and 40 weeks. Exclusion criteria comprised pre-existing diabetes mellitus (Type 1 or 2), history of prior urinary tract surgery or known urogenital anomalies, current urinary tract infection (confirmed by urine analysis/culture), known neurological conditions affecting bladder function, multiple pregnancies, history of psychiatric illness, use of medications known to affect bladder function, and refusal to participation to study.

Data collection and definitions

GDM was diagnosed according to either the two-step approach (American College of Obstetricians and Gynecologists [ACOG] guidelines: 50-g glucose challenge test followed by 100-g oral glucose tolerance test [OGTT] for screen-positive individuals) or the one-step criteria (International Association of the Diabetes and Pregnancy Study Groups [IADPSG]: 50-g OGTT)[9,10].

Sociodemographic information included age, parity, educational level, occupation, income, smoking, alcohol consumption, coffee/tea consumption, previous mode of delivery, and daytime/nighttime urination frequency. Gestational weight gain was calculated as the difference between current weight and pre-pregnancy weight.

Urogenital Distress Inventory-6

This 6-item instrument measures distress from irritative (items 1–2), stress (items 3–4), and obstructive/voiding symptoms (items 5–6) on a 4-point Likert scale (0="not at all" to 3="greatly"). Higher total scores indicate greater distress[11]. The Turkish version was validated by Çam and Sakallı, showing strong reliability (Cronbach's α=0.74–0.87)[12].

Incontinence Impact Questionnaire-7

This 7-item scale evaluates UI's impact on physical activity, social relationships, travel, and emotional well-being using a 4-point Likert scale. Scores are standardized (0–100); higher scores denote greater impact. The Turkish version demonstrated excellent internal consistency (Cronbach's α=0.87) and test–retest reliability (Spearman's ρ=0.99)[12].

International Consultation on Incontinence Questionnaire-Short Form

This instrument assesses the severity and impact of UI over the past 4 weeks. The total score is derived from three core items evaluating frequency, volume of leakage, and the impact on QoL (items 3–5), yielding a 0–21 severity score, with higher scores indicating worse symptoms. Items 1 and 2 are demographic, while item 6 assesses the perceived cause of leakage and is not scored. The Turkish version of the ICIQ-SF has been validated by Cetinel et al[13].

Sandvik Severity Index (Incontinence Severity Index)

This index quantifies UI severity by multiplying the frequency of leakage episodes by the amount of leakage. Based on the resulting score, UI is classified as mild (1–2), moderate (3–6), or severe (8–12). The ISI has been widely used in clinical research to stratify incontinence severity[14].

Statistical analysis

Analyses were performed using IBM® SPSS® Statistics 30. Continuous variables are summarized as mean±standard deviation or median (interquartile range) based on normality assessed via Kolmogorov-Smirnov test, skewness, and kurtosis. Categorical variables are presented as frequencies (percentages). Non-parametric tests were employed due to non-normal distributions. Group comparisons (GDM vs. control) for UDI-6, ICIQ-SF, IIQ-7, and Sandvik Severity Index scores used the Mann-Whitney U test (effect size: $r=Z/\sqrt{N}$; thresholds: 0.10=small, 0.30=medium, and 0.50=large). Categorical variable homogeneity was assessed using the chi-square

test. Relationships between continuous variables were explored via Spearman's correlation analysis. Multivariable logistic regression was performed to examine the independent association between GDM and UI (defined as ICIQ-SF score ≥1) while adjusting for potential confounding variables including age, pre-pregnancy body mass index (BMI), BMI during pregnancy, and gestational weight gain. Adjusted odds ratios (ORs) with 95%CIs were calculated. Questionnaire internal consistency was evaluated using Cronbach's alpha. Statistical significance was defined as p<0.05.

RESULTS

A total of 500 pregnant women participated in the study, with 251 in the GDM group and 249 in the control (non-GDM) group. The sociodemographic characteristics of the participants are presented in Table 1.

The mean age of the participants with GDM was significantly higher than that of the control group (30.96±6.23 vs. 27.96±5.02 years; p<0.001). Pre-pregnancy BMI and BMI during pregnancy were also significantly higher among those with GDM (p<0.001). The median gestational weight gain was significantly greater in the GDM group (15 kg vs. 14 kg; p=0.027). No significant differences were found in terms of smoking, parity, education level, employment status, income level, alcohol consumption, coffee and tea intake, previous mode of delivery, and frequency of nighttime urination (p>0.05) (Tables 1 and 2).

Pregnant women with GDM scored significantly higher across all incontinence and QoL instruments compared to those without GDM (p<0.001 for all comparisons, Table 3). The mean ICIQ-SF score was significantly higher in the GDM group (7.16±5.65) than that in the control group (4.55±5.37; p<0.001). The UDI-6 scores were also significantly elevated in the GDM group (8.87±4.41 vs. 6.83±4.52; p<0.001). The IIQ-7 and ISI scores were higher in the GDM group (7.39±7.13 vs. 4.45±6.34, p<0.001; and 2.81±2.97 vs. 1.74±2.73; p<0.001, respectively) (Table 3).

Multivariable logistic regression analysis was conducted to assess whether GDM remained independently associated with

Table 1. Sociodemographic characteristics of pregnant women with and without gestational diabetes mellitus.

Groups	GDM group (n=251)	Control group (n=249)	Total (n=500)	Test value	p-value
Age (years)	30.96±6.2330 (26–35)	27.96±5.0228 (24–31)	29.47±5.8529 (25–33)	Z=-5.062	**p<0.001**
Cigarettes/day	0.92±3.250 (0–0)	0.43±1.550 (0–0)	0.68±2.560 (0–0)	Z=-1.001	p=0.317
Education level					
Primary school	40 (15.9%)	33 (13.3%)	73 (14.6%)	χ^2=1.197	p=0.754
Middle school	35 (13.9%)	38 (15.3%)	73 (14.6%)		
High school	127 (50.6%)	123 (49.4%)	250 (50.0%)		
University	49 (19.5%)	55 (22.1%)	104 (20.8%)		
Employment status					
Employed	47 (18.7%)	46 (18.5%)	93 (18.6%)	χ^2=0.000	p=1.000
Unemployed	204 (81.3%)	203 (81.5%)	407 (81.4%)		
Income status					
Income<expenses	15 (6.0%)	21 (8.4%)	36 (7.2%)	χ^2=2.095	p=0.351
Income=expenses	203 (80.9%)	203 (81.5%)	406 (81.2%)		
Income>expenses	33 (13.1%)	25 (10.0%)	58 (11.6%)		
Smoking status					
Yes	29 (11.6%)	22 (8.8%)	51 (10.2%)	χ^2=0.734	p=0.392
No	222 (88.4%)	227 (91.2%)	449 (89.8%)		
Coffee consumption					
0 cups/day	139 (55.4%)	123 (49.4%)	262 (52.4%)	χ^2=1.561	p=0.212
1–3 cups/day	112 (44.6%)	126 (50.6%)	238 (47.6%)		
Tea consumption					
0 cups/day	34 (13.5%)	29 (11.6%)	63 (12.6%)	χ^2=3.993	p=0.136
1–3 cups/day	127 (50.6%)	148 (59.4%)	275 (55.0%)		
≥4 cups/day	90 (35.9%)	72 (28.9%)	162 (32.4%)		

Data are presented as mean±standard deviation or median (interquartile range) for continuous variables and number (percentage) for categorical variables. $\bar{x}$: mean; SD: standard deviation; M (IQR): median (25th–75th percentiles); n: number; %: percent; Z: Mann-Whitney U standardized test statistic; χ^2: chi-square test; GDM: gestational diabetes mellitus. Statistically significant values are denoted in bold.

Table 2. Distribution of sociodemographic, obstetric, and incontinence-related characteristics of pregnant women by group.

Group	GDM group (n=251)	Control group (n=249)	Total (n=500)	Test value	p-value
Pre-pregnancy BMI (kg/m²)	26.20±3.4726 (24–28)	24.39±3.8524 (22–26)	25.30±3.7725 (23–27)	Z=-6.090	**p<0.001**
BMI during pregnancy (kg/m²)	31.93±3.6432 (30–34)	29.79±4.0929 (27–32)	30.86±4.0131 (29–33)	Z=-6.917	**p<0.001**
Gestational weight gain (kg)	15.04±5.4715 (12–18)	14.25±4.6314 (11–17)	14.65±5.0814 (11–18)	Z=-2.205	**p=0.027**
Gestational week	38.60±1.6439 (38–40)	38.63±1.8039 (38–40)	38.62±1.7239 (38–40)	Z=-0.807	p=0.420
Number of pregnancies	2.27±1.292 (1–3)	2.23±1.332 (1–3)	2.25±1.312 (1–3)	Z=-0.617	p=0.537
Number of living children	0.98±1.021 (0–2)	0.92±1.071 (0–2)	0.95±1.041 (0–2)	Z=-1.107	p=0.268
Previous birth type				χ^2=3.167[a]	p=0.205
Vaginal	97 (38.6%)	114 (45.8%)	211 (42.2%)		
Cesarean	59 (23.5%)	46 (18.5%)	105 (21.0%)		
Instrumental (vacuum/kiwi)	95 (37.8%)	89 (35.7%)	184 (36.8%)		
Nighttime urination frequency				χ^2=6.025[a]	p=0.197
Never	5 (2.0%)	2 (0.8%)	7 (1.4%)		
Once	23 (9.2%)	28 (11.2%)	51 (10.2%)		
Twice	65 (25.9%)	84 (33.7%)	149 (29.8%)		
Three times	78 (31.1%)	68 (27.3%)	146 (29.2%)		
Four times	80 (31.9%)	67 (26.9%)	147 (29.4%)		

Data are presented as mean±standard deviation or median (interquartile range) for continuous variables and number (percentage) for categorical variables. x̄ mean; SD: standard deviation; M (IQR): median (25th–75th percentiles); n: number; %: percent; Z: Mann-Whitney U standardized test statistic; χ^2: chi-square test; GDM: gestational diabetes mellitus; BMI: body mass index. [a]Pearson's chi-square test. Statistically significant values are denoted in bold.

Table 3. Comparison of Incontinence Severity Index, Urogenital Distress Inventory-6, Incontinence Impact Questionnaire-7, and International Consultation on Incontinence Questionnaire-Short Form scores by group and reliability coefficients

Scale	GDM group (n=251)	Control group (n=249)	Z	p-value	r	Cronbach's α (total)
ICIQ-SF	7.16±5.65 (0–21)	4.55±5.37 (0–21)	-5.022	<0.001	0.22	0.855
UDI-6	8.87±4.41 (0–18)	6.83±4.52 (0–16)	-5.334	<0.001	0.24	0.793
IIQ-7	7.39±7.13 (0–21)	4.45±6.34 (0–21)	-4.898	<0.001	0.22	0.955
ISI	2.81±2.97 (0–9)	1.74±2.73 (0–9)	-5.838	<0.001	0.26	0.917

M (IQR): median (interquartile range: 25th–75th percentile); x̄: mean; SD: standard deviation; min–max: minimum and maximum values; α: Cronbach's alpha reliability coefficient; Z: standardized Z value from Mann-Whitney U test; r: effect size for Mann-Whitney U test; GDM: gestational diabetes mellitus; ICIQ-SF: International Consultation on Incontinence Questionnaire-Short Form; UDI-6: Urogenital Distress Inventory-6; IIQ-7: Incontinence Impact Questionnaire-7; ISI: Incontinence Severity Index

UI after controlling for known confounding variables. In the unadjusted model, GDM was significantly associated with UI (OR 2.87, 95%CI 1.96–4.21, p<0.001). After adjusting for age, pre-pregnancy BMI, BMI during pregnancy, and gestational weight gain, GDM remained independently and significantly associated with increased UI risk (adjusted OR 2.34, 95%CI 1.56–3.51, p<0.001).

DISCUSSION

This study shows that GDM is associated with significantly higher prevalence and severity of LUTS, including UI, as well as a greater negative impact on incontinence-related QoL among pregnant women. The GDM group scored significantly higher on all validated questionnaires (ICIQ-SF, UDI-6, IIQ-7, and ISI) designed to assess UI frequency, amount, symptom bother (urogenital distress), and impact on daily life. In addition, multivariable regression analysis demonstrated that GDM remained independently associated with increased UI even after adjusting for age, BMI, and gestational weight gain. Our findings demonstrate consistency with large-scale studies from North America, Europe, and Asia that have reported elevated LUTS/UI prevalence in pregnant women with GDM compared to normoglycemic controls[15-18].

Several potential pathophysiological mechanisms could explain the observed relationship between GDM and increased LUTS/UI during pregnancy. Chronic hyperglycemia, a hallmark of diabetes and present to varying degrees in GDM, is known to adversely affect various tissues, including muscle, nerve, and connective tissue, all of which are critical for maintaining pelvic floor integrity and function[16]. Diabetic neuropathy could impair the neural control of the bladder detrusor muscle and urethral sphincters, potentially leading to symptoms such as urgency, frequency, and incomplete emptying, as well as sphincter weakness contributing to stress incontinence[6,17]. Diabetic myopathy might weaken the PFMs, which are crucial for urethral support and continence. Furthermore, diabetes can disrupt connective tissue metabolism, affecting the collagen and elastin content and organization in pelvic support structures, potentially leading to fascial and ligamentous laxity[16,18].

Our multivariable regression analysis specifically addressed the important consideration raised regarding the influence of factors that were more prevalent in the GDM group, including age and BMI. The GDM participants were significantly older and had higher pre-pregnancy and pregnancy BMIs, as well as greater gestational weight gain. These factors are well-established independent risk factors for LUTS and UI during and after pregnancy[19]. Higher BMI and greater weight gain increase intra-abdominal pressure, which can strain the pelvic floor and exacerbate incontinence. However, the persistence of a significant GDM–UI association supports the notion that the metabolic derangements specific to GDM, including chronic hyperglycemia, oxidative stress, and inflammatory cytokine dysregulation, may directly affect pelvic floor tissues through mechanisms distinct from those related to mechanical factors alone[8,15].

The significantly higher scores on the UDI-6 and IIQ-7 questionnaires in the GDM group are particularly concerning, highlighting the substantial negative impact of LUTS/UI on their QoL. UI and bothersome LUTS can lead to social isolation, avoidance of physical activity, emotional distress, anxiety, depression, and reduced self-confidence during a period that already presents unique physical and psychological challenges[3,20]. Qualitative investigations from various cultural contexts have documented that pregnant women with GDM experience compounded distress when LUTS/UI symptoms are superimposed upon the already-demanding self-management requirements of glycemic control, dietary modification, and frequent monitoring[21,22]. This intersection of metabolic and urogenital symptom burden underscores the imperative for integrated, holistic management approaches.

These findings have significant implications for clinical practice. Routine, systematic screening for LUTS/UI may be considered for all pregnant women with GDM. Brief, validated instruments such as the ICIQ-SF can be efficiently integrated into standard antenatal care. Early identification of symptomatic women facilitates timely, conservative interventions, particularly supervised PFM training, which is effective in reducing pregnancy-associated incontinence[23,24]. Recent randomized trials demonstrate significant continence and QoL improvements with pelvic floor interventions specifically in pregnant women with GDM[16-20]. Pregnant women with GDM and moderate-to-severe LUTS/UI may benefit from referral to urogynecology or pelvic floor physical therapy for comprehensive assessment and individualized treatment. Proactive management of LUTS/UI in pregnant women with GDM may improve immediate QoL and potentially reduce persistent postpartum pelvic floor dysfunction, conferring benefits beyond the pregnancy.

Limitations

While this study provides strong evidence of the association between GDM and increased LUTS/UI and diminished QoL during pregnancy, it is not without limitations. As a case–control study, our design cannot definitively establish causal relationships or determine the temporal sequence of GDM diagnosis and LUTS/UI onset. Longitudinal studies are needed to track the incidence and progression of LUTS/UI throughout pregnancy and into the postpartum period in women with GDM compared to non-GDM controls. Prospective cohort investigations that assess women prior to GDM diagnosis and follow them longitudinally would provide stronger evidence regarding causality and enable examination of dose–response relationships between glycemic control and symptom severity. Investigating the efficacy of specific PFM training or other conservative management programs specifically in the GDM population is also warranted. Although validated questionnaires are robust tools for assessing subjective symptoms and impact, complementing these with objective measures, such as bladder diaries, pad tests, or urodynamic studies, where clinically indicated, could provide a more complete picture of bladder function. Finally, while we attempted to match groups by gestational age, we acknowledge that the GDM and control groups were not fully homogeneous with respect to age and BMI. Although we addressed this limitation through multivariable regression adjustment, future case–control studies employing individual matching or propensity score matching on these variables would further strengthen causal inference.

CONCLUSION

This study provides compelling evidence that GDM is significantly associated with increased severity of LUTS/UI. Women with GDM reported higher levels of UI severity, symptom distress, and impact on QoL compared to normoglycemic pregnant women.

ETHICAL STATEMENT

Written informed consent was sought from each pregnant woman following an explanation of the purpose and design of the study. Pregnant women were informed that their involvement in the study was voluntary and that they could withdraw at any time. In addition, they were given the assurance that their data would be kept confidential. This study followed the ethical principles outlined in the Declaration of Helsinki and was approved by the Research Ethics Committee (approval: 49, 5.04.2023).

AUTHORS' CONTRIBUTIONS

LK: Conceptualization, Data curation, Writing – original draft, Writing – review & editing. **EK:** Conceptualization, Writing – original draft, Writing – review & editing. **ZK:** Data curation, Writing – original draft, Writing – review & editing. **HD:** Formal Analysis, Writing – original draft, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Metsärinne K, Pietilä M, Kantola I, Stenman LK, Hätinen OP, Vesikansa A, et al. The majority of type 2 diabetic patients in Finnish primary care are at very high risk of cardiovascular events: a cross-sectional chart review study (STONE HF). Prim Care Diabetes. 2022;16(1):135-41. https://doi.org/10.1016/j.pcd.2021.12.012
2. Simmons D, Immanuel J, Hague WM, Teede H, Nolan CJ, Peek MJ, et al. Treatment of gestational diabetes mellitus diagnosed early in pregnancy. N Engl J Med. 2023;388(23):2132-44. https://doi.org/10.1056/NEJMoa2214956
3. Craig L, Sims R, Glasziou P, Thomas R. Women's experiences of a diagnosis of gestational diabetes mellitus: a systematic review. BMC Pregnancy Childbirth. 2020;20(1):76. https://doi.org/10.1186/s12884-020-2745-1
4. Tähtinen RM, Cartwright R, Tsui JF, Aaltonen RL, Aoki Y, Cárdenas JL, et al. Long-term impact of mode of delivery on stress urinary incontinence and urgency urinary incontinence: a systematic review and meta-analysis. Eur Urol. 2016;70(1):148-58. https://doi.org/10.1016/j.eururo.2016.01.037
5. Løwenstein E, Jepsen R, Andersen LL, Laigaard J, Møller LA, Gaede P, et al. Prevalence of urinary incontinence among women with diabetes in the Lolland-Falster Health Study, Denmark. Neurourol Urodyn. 2021;40(3):855-67. https://doi.org/10.1002/nau.24636
6. Vesentini G, Barbosa AMP, Damasceno DC, Marini G, Piculo F, Matheus SMM, et al. Alterations in the structural characteristics of rectus abdominis muscles caused by diabetes and pregnancy: a comparative study of the rat model and women. PLoS One. 2020;15(4):e0231096. https://doi.org/10.1371/journal.pone.0231096
7. Tagami K, Iwama N, Hamada H, Tomita H, Kudo R, Kumagai N, et al. Advanced maternal age is a risk factor for both early and late gestational diabetes mellitus: the Japan Environment and Children's Study. J Diabetes Investig. 2025;16(4):735-43. https://doi.org/10.1111/jdi.14400
8. Sartorao Filho CI, Nunes SK, Magyori ABM, Calderon IMP, Barbosa AMP, Rudge MVC. The role of gestational diabetes mellitus and pelvic floor 3D-ultrasound assessment during pregnancy predicting urinary incontinence: a prospective cohort study. BMC Pregnancy Childbirth. 2023;23(1):637. https://doi.org/10.1186/s12884-023-05932-8
9. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2011;34(Suppl 1):S62-9. https://doi.org/10.2337/dc11-S062
10. International Association of Diabetes and Pregnancy Study Groups Consensus Panel, Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care. 2010;33(3):676-82. https://doi.org/10.2337/dc09-1848
11. Uebersax JS, Wyman JF, Shumaker SA, McClish DK, Fantl JA. Short forms to assess life quality and symptom distress for urinary incontinence in women: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. Continence Program for Women Research Group. Neurourol Urodyn. 1995;14(2):131-9. https://doi.org/10.1002/nau.1930140206
12. Cam C, Sakalli M, Ay P, Cam M, Karateke A. Validation of the short forms of the incontinence impact questionnaire (IIQ-7) and the urogenital distress inventory (UDI-6) in a Turkish population. Neurourol Urodyn. 2007;26(1):129-33. https://doi.org/10.1002/nau.20292
13. Cetinel B, Ozkan B, Can G. The validation study of ICIQ-SF Turkish version. Turk Uroloji Dergisi. 2004;30:332-8.
14. Hazar U, Şirin A. İnkontinans şiddet indeksinin geçerlik ve güvenirliği çalışması. Adnan Menderes Üniv Tıp Fakültesi Dergisi. 2008;9:5-8.
15. Cozier YC, Castro-Webb N, Bertrand KA, Ndama M, Chai TC, Harlow BL, et al. Gestational and Type 2 diabetes in relation to urinary incontinence in black women in the U.S. *Uro.* 2025;5(2):8. https://doi.org/10.3390/uro5020008
16. Wu Y, Li T, Cai F, Ye X, Xu M. Stable pelvic floor muscle training improves urinary incontinence in women with gestational diabetes mellitus. J Obstet Gynaecol. 2024;44(1):2420192. https://doi.org/10.1080/01443615.2024.2420192
17. França DCH, França EL, Sobrevia L, Barbosa AMP, Honorio-França AC, Rudge MVC, et al. Integration of nutrigenomics, melatonin, serotonin and inflammatory cytokines in the pathophysiology of pregnancy-specific urinary incontinence in women with gestational diabetes mellitus. Biochim Biophys Acta Mol Basis Dis. 2023;1869(6):166737. https://doi.org/10.1016/j.bbadis.2023.166737

18. Arslan M, Kozan R. Pelvic floor dysfunction in patients with gestational diabetes mellitus. World J Diabetes. 2025;16(2):99823. https://doi.org/10.4239/wjd.v16.i2.99823
19. Ansarzadeh S, Salehi L, Mahmoodi Z, Mohammadbeigi A. Factors affecting the quality of life in women with gestational diabetes mellitus: a path analysis model. Health Qual Life Outcomes. 2020;18(1):31. https://doi.org/10.1186/s12955-020-01293-4
20. Yavuz A, Kocaöz S, Kara P, Destegül E. The effects of gestational diabetes on lower urinary tract symptoms of pregnant women: a case-control study. J Obstet Gynaecol. 2022;42(8):3531-6. https://doi.org/10.1080/01443615.2022.2152657
21. Henderson J, Alderdice F, Redshaw M. Factors associated with maternal postpartum fatigue: an observationalstudy. BMJ Open. 2019;9(7):e025927. https://doi.org/10.1136/bmjopen-2018-025927
22. Parsons J, Sparrow K, Ismail K, Hunt K, Rogers H, Forbes A. A qualitative study exploring women's health behaviours after a pregnancy with gestational diabetes to inform the development of a diabetes prevention strategy. Diabet Med. 2019;36(2):203-13. https://doi.org/10.1111/dme.13794
23. Mørkved S, Bø K. Effect of pelvic floor muscle training during pregnancy and after childbirth on prevention and treatment of urinary incontinence: a systematic review. Br J Sports Med. 2014;48(4):299-310. https://doi.org/10.1136/bjsports-2012-091758
24. Woodley SJ, Boyle R, Cody JD, Mørkved S, Hay-Smith EJC. Pelvic floor muscle training for prevention and treatment of urinary and faecal incontinence in antenatal and postnatal women. Cochrane Database Syst Rev. 2017;12(12):CD007471. https://doi.org/10.1002/14651858.CD007471.pub3

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251488

A comparative evaluation of ChatGPT and Gemini for stroke caregiver support: accuracy, safety, and response characteristics

Osman Karaca[1]*, Mustafa Savaş Torlak[2], Ayşe Nur Tekin[3]

SUMMARY

OBJECTIVE: The use of large language models is increasing in healthcare. The aim of this study was to evaluate and compare the responses provided by ChatGPT (OpenAI) and Gemini (Google) to frequently asked questions about stroke and its treatment by caregivers, in terms of clarity, empathy/tone, actionability, information risk/safety, and accuracy.

METHODS: This study used a qualitative content analysis to evaluate responses generated by large language models to stroke-related caregiving questions. Ten frequently asked questions were identified through a review of the literature on stroke caregiving. These questions were presented to the free versions of ChatGPT and Gemini, and responses from each model were recorded. The recorded responses were independently evaluated by 10 experts from neurology, physiatry, and physiotherapy using a structured Likert-scale instrument across five predefined categories: clarity, empathy/tone, actionability, information risk/safety, and accuracy.

RESULTS: Generalized estimating equation analyses indicated that the overall trends in response scores did not differ significantly between ChatGPT and Gemini across all evaluated categories ($p>0.05$). In contrast, response score trends varied primarily in clarity and empathy/tone, with limited variation in actionability ($p<0.05$). Intraclass correlation coefficient values were below 0.50 for both ChatGPT and Gemini.

CONCLUSION: Although both models generally performed well, question-level variations and low inter-rater reliability limit generalizability, and even small shortcomings in information risk may have important implications for stroke caregivers and, indirectly, for patients with stroke.

KEYWORDS: Artificial intelligence. Stroke rehabilitation. Natural language processing. Information sources. Patient safety.

INTRODUCTION

Natural language processing enables artificial intelligence (AI) systems to analyze text and generate or interpret human-like language[1]. In this context, ChatGPT (OpenAI) and Gemini (Google) are prominent examples of large language models (LLMs), implemented as AI-powered, multimodal systems[2]. If some ethical and technical issues are carefully addressed, they have the potential to serve as valuable tools for improving healthcare[3].

Stroke was reported as a leading cause of death and disability, making both stroke prevention and post-stroke rehabilitation critically important[4,5]. After a stroke, harmony between the patient, their relatives, caregivers, and healthcare professionals plays a role in recovery[6]. The post-stroke period is crucial, and guidance from professionals helps patients and caregivers cope with stroke outcomes[7]. Heavy caregiving demands can lead caregivers to use escape or distancing strategies; tailored programs are needed to foster effective coping and protect their well-being[8]. Adequate stroke care education is essential for caregivers and may help reduce uncertainty and knowledge gaps by addressing informational and psychosocial needs[9]. Post-stroke caregivers have diverse care-related needs, while hospital-provided information is often limited or outdated; consequently, caregivers increasingly seek health information online[10].

There is information in the literature that the use of LLMs on the Internet is increasing, and they can contribute as a facilitating factor in the rehabilitation of neurological diseases[11]. It is also stated that such programs can be beneficial for health professionals[12,13], and relatives of patients[14]. In recent years, there has been a growing number of studies comparing the accuracy and effectiveness of ChatGPT and Gemini in responding to medical questions[15].

Despite widespread use in healthcare, evidence on AI tools' ability to address the psychosocial and practical needs of stroke caregivers is limited. This study aimed to evaluate and compare responses generated by ChatGPT and Gemini to frequently asked stroke-related caregiving questions in terms of clarity, empathy/tone, actionability, information risk/safety, and accuracy.

[1]KTO Karatay University, Department of Physiotherapy and Rehabilitation – Konya, Türkiye.
[2]KTO Karatay University, Vocational School of Health Sciences, Departments of Therapy and Rehabilitation – Konya, Türkiye.
[3]Baskent University Hospital Konya, Department of Physical Therapy and Rehabilitation – Konya, Türkiye.
*Corresponding author: osman.karaca@karatay.edu.tr
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on January 14, 2026. Accepted on January 17, 2026.
Scientific Editor: José Maria Soares Júnior

METHODS

This study was conducted in December 2025. At the beginning of the study, browser histories were cleared, and new accounts were created for both ChatGPT and Gemini to minimize the influence of prior interactions.

Based on a literature review[16-20], the following 10 questions reflecting key stroke caregiver needs, including risk reduction, recovery and rehabilitation, psychosocial support, home care, and social and administrative resources, were presented to both ChatGPT and Gemini, and the generated responses were collected for analysis; full responses are available from the corresponding author upon request.

1. How can we reduce the risk of having another stroke?
2. What is the likelihood of regaining the functions lost due to stroke?
3. How long does the recovery process take after a stroke?
4. How can I support a stroke survivor experiencing mood changes or depression?
5. How can I motivate a stroke survivor who has lost motivation for rehabilitation?
6. How can I prevent caregiver burnout?
7. What treatment and therapy options are available for stroke survivors?
8. How can I access physiotherapy services for a stroke survivor, and how can I tell if it is effective?
9. How can I properly assist a stroke survivor at home?
10. What disability pension or social support benefits are provided by the government for stroke survivors, and how can I apply for them?

Responses from ChatGPT and Gemini were independently evaluated by 10 blinded experts (three neurologists, three physiatrists, and four physiotherapists), each with ≥10 years of professional experience.

In the absence of a standardized assessment tool for AI-generated health information, evaluation domains were informed by established frameworks used to evaluate patient-facing health information and AI-generated medical content, emphasizing clarity, actionability, tone and empathy, potential information risk/safety, and accuracy[20-23]. Five domains were assessed using structured Likert scales: clarity, empathy/tone, actionability, and information risk/safety (all scored from 1 to 5, with higher scores indicating better performance), and accuracy (scored from 1 to 4, with higher scores indicating greater accuracy; the original scoring was revised for consistency)[22]. Clarity was defined as the extent to which the response was easily understandable by the target reader; empathy/tone as the degree to which the response addressed the emotional and psychosocial needs of stroke caregivers; actionability as the extent to which the response provided clear, concrete, and feasible recommendations that caregivers could implement; information risk/safety as the extent to which the response included potentially unsafe or misleading information; and accuracy as the degree to which the response was correct and sufficiently comprehensive to meet the caregiver's needs without requiring additional clarification or containing misleading or incorrect information. Experts received a written guideline explaining the evaluation domains and Likert-scale anchors prior to rating.

Statistical analysis

All analyses were conducted using IBM SPSS Statistics version 22. Ordinal data are presented as medians and interquartile ranges (25th–75th percentiles). To account for correlated ratings from the same expert (n=10), separate ordinal generalized estimating equation (GEE) models were fitted for each evaluation domain (clarity, empathy/tone, actionability, information risk/safety, and accuracy), using a multinomial distribution with a cumulative logit link. Evaluator was specified as the subject variable, and model type and question ID were included as within-subject factors, with an exchangeable correlation structure.

For the information risk and accuracy domains, full-scale ordinal GEE models did not converge reliably; therefore, scores were dichotomized for sensitivity analyses (Information Risk/Safety: [4–5] vs. [1–3]; Accuracy: [3–4] vs. [1–2]). Inter-rater reliability was assessed using the intraclass correlation coefficient (ICC). According to established benchmarks, ICC values <0.50 indicate poor reliability[24]. A $p<0.05$ was considered statistically significant.

Ethical approval was obtained from the KTO Karatay University Ethics Committee (Approval No. 2025/038, 8 May 2025), and the study was conducted in accordance with the Declaration of Helsinki and national ethical standards. All participants provided written informed consent.

RESULTS

This qualitative content analysis evaluated responses generated by ChatGPT and Gemini to frequently asked stroke-related caregiving questions. Median and interquartile range (IQR [25th–75th percentiles]) of the scores for the responses to each question are presented in Table 1. Median values are calculated across all categories for each question.

Median and IQR of the scores for each category are presented in Table 2. Median values are calculated across all questions for each category.

In the category-level analysis, no statistically significant differences were observed between the two models for all evaluated categories (p>0.05) (Table 2).

Results of ordinal GEE analyses are shown for each question, combining data from all models. At question-level analyses, Q1, Q6, and Q9 in Clarity; Q4, Q5, and Q6 in empathy/tone; and Q9 in Actionability showed moderately higher scores compared with the reference question (p<0.05). In empathy/tone, Q9 showed borderline significance (p=0.05). Answers for the other questions did not differ significantly (p>0.05) (Table 3).

Differences were most apparent in questions addressing psychological concerns and home-based patient assistance, which were more likely to receive higher evaluation scores. This pattern may reflect question-specific characteristics, as such questions inherently involve greater emotional and psychosocial components. However, ICC values were below 0.50 for both ChatGPT and Gemini, indicating poor inter-rater reliability.

DISCUSSION

This study aimed to evaluate and compare the responses provided by ChatGPT and Gemini to frequently asked questions about stroke, its treatment, and caregiver-related concerns, in terms of clarity, empathy/tone, actionability, information risk/safety, and accuracy. Results indicated that the overall trends in response scores did not differ significantly between the language models across all evaluated categories. Response score trends differed significantly for some individual questions.

In a review study investigating the adequacy of LLMs in patient information, it was stated that LLMs have significant potential in patient information, but they contain deficiencies in terms of accuracy and readability, and also contain ethical concerns[21]. Another study evaluated five stroke-related questions on ChatGPT 3.5; while answers were generally adequate, sources were often inaccurate, indicating

Table 1. Overall median (interquartile range) scores across all five evaluation categories for each question.

	ChatGPT		Gemini	
	Median	IQR	Median	IQR
Q1	4.00	3.00–4.00	4.00	3.00–4.00
Q2	4.00	3.00–4.00	3.00	3.00–4.00
Q3	4.00	3.00–4.00	3.00	3.00–4.00
Q4	4.00	3.00–5.00	4.00	3.00–5.00
Q5	4.00	3.00–4.00	4.00	3.00–5.00
Q6	4.00	3.00–5.00	3.00	3.00–5.00
Q7	3.00	3.00–4.00	4.00	3.00–4.00
Q8	4.00	3.00–4.00	4.00	3.00–4.00
Q9	4.00	3.00–4.00	4.00	3.00–5.00
Q10	3.50	3.00–4.00	4.00	3.00–4.00

Median values are calculated across all categories for each question. Q: Question; IQR: interquartile range (25th–75th percentiles).

Table 2. Median (interquartile range) scores and ordinal generalized estimating equation estimates by evaluation category.

Category	ChatGPT	Gemini	B (95%CI)	Std. error	Wald χ^2	p
	Median IQR	Median IQR				
Clarity (1–5)	4.00 4.00–5.00	4.00 3.00–5.00	0.25 (-0.26, 0.76)	0.53	0.91	0.34
Empathy/ tone (1–5)	4.00 3.00–4.00	4.00 3.00–4.00	0.13 (-0.33, 0.59)	0.26	0.29	0.59
Actionability (1–5)	4.00 3.00–4.00	4.00 3.00–4.00	0.15 (-0.30, 0.61)	0.23	0.43	0.21
Information risk (1–5)	4.00 3.00–4.00	4.00 3.00–4.00	0.14 (-0.48, 0.76)	0.31	0.20	0.65[S]
Accuracy (1–4)	3.00 2.00–3.00	3.00 2.00–3.00	0.15 (-0.50, 0.79)	0.33	0.20	0.65[S]

IQR: interquartile range (25th–75th percentiles); B: parameter estimate; CI: confidence interval; [S]p-value derived from GEE sensitivity analysis with dichotomized scores, Gemini was used as the reference model in the ordinal GEE analyses.

Table 3. Ordinal generalized estimating equation estimates results by question.

	Clarity		Empathy/tone		Actionability		Information risk		Accuracy	
	B	p	B	p	B	p	B	p^s	B	p^s
Q1	1.07	**0.04**	0.24	0.38	0.35	0.44	0.00	1.00	-0.21	0.81
Q2	0.17	0.77	0.50	0.13	0.47	0.33	-0.21	0.70	-0.62	0.43
Q3	0.26	0.66	0.29	0.41	0.17	0.67	0.48	0.40	-0.42	0.56
Q4	1.32	0.28	0.87	**0.006**	0.53	0.12	0.77	0.15	0.48	0.48
Q5	0.73	0.11	0.70	**0.009**	0.68	0.11	0.23	0.65	0.23	0.74
Q6	1.08	**0.046**	0.81	**0.01**	0.49	0.31	0.21	0.74	1.11	0.10
Q7	0.26	0.47	0.11	0.46	0.09	0.72	0.00	1.00	-0.62	0.30
Q8	0.73	0.86	0.53	0.13	0.39	0.31	0.23	0.56	0.77	0.40
Q9	1.37	**0.02**	0.51	0.050*	0.79	**0.03**	0.48	0.32	1.11	0.14
Q10	Reference									

Q: question, B: parameter estimate, ^{s}p-value derived from GEE sensitivity analysis with dichotomized scores, p*: borderline significance. Statistically significant values are denoted in bold.

caution is needed[23]. Only one prior study (Neo et al.) has evaluated AI chatbot responses to questions from stroke patients and caregivers, reporting safety-, accuracy-, and relevance-related concerns despite broadly similar model performance[20]. Consistent with these findings, our study identified information risk/safety as a key limitation of AI-generated health advice, particularly for questions requiring symptom-specific safety considerations. Unlike Neo et al., who included both patients and caregivers, our study focused exclusively on caregivers, extending these safety-related concerns to caregiver-facing use. Qualitative examination of responses with lower information risk/safety scores revealed several recurring limitations. A prominent issue was the lack of personalization in the recommendations. For instance, in responses addressing daily care strategies for individuals with stroke, cognitive impairments commonly observed after stroke were insufficiently considered, potentially limiting the applicability of the advice. In other cases, recommendations implicitly assumed access to specific equipment or materials that may not be available to all caregivers. Experts noted that such assumptions could not only reduce feasibility but also inadvertently contribute to feelings of inadequacy or failure among caregivers when the suggested resources are unavailable. Overall, while both models showed similar scoring patterns, lower scores in the information risk domain suggest that some responses may still carry potential adverse consequences.

This study has several limitations. First, the findings represent a static snapshot of AI performance, as LLMs are updated frequently and responses may change over time. Second, ratings were based on expert judgment rather than comparison with a single verified clinical guideline. The low inter-rater agreement observed in this study is an important limitation and likely reflects the subjective nature of multidisciplinary expert evaluations, a challenge also reported in prior assessments of AI-generated health content[20]. Lastly, responses were evaluated by experts rather than stroke caregivers themselves; perceptions of empathy, clarity, and actionability may therefore differ from real end-user experiences. Future research should prioritize user-centered evaluations involving caregivers, the development of AI systems informed by curated caregiver-specific resources, and the exploration of hybrid models combining AI outputs with human moderation for sensitive caregiving guidance.

CONCLUSION

Both models generally received favorable ratings across all evaluated categories and performed similarly across these domains. However, variations at the question level suggest that question content can influence response scores. Therefore, the findings should be interpreted with caution, and generalization of these results across different caregiving questions may not be appropriate. Notably, some lower scores observed in the information risk/safety category suggest that some responses may carry potential risks. In this context, the use of LLMs by caregivers of individuals with stroke may have the potential to lead to unintended adverse consequences if relied upon without appropriate clinical guidance.

AUTHORS' CONTRIBUTIONS

OK: Data curation, Methodology, Writing – original draft. **MST:** Data curation, Formal Analysis, Methodology, Writing – review & editing. **ANT:** Data curation, Methodology, Writing – original draft.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Khurana D, Koli A, Khatter K, Singh S. Natural language processing: state of the art, current trends and challenges. Multimed Tools Appl. 2023;82(3):3713-44. https://doi.org/10.1007/s11042-022-13428-4
2. Imran M, Almusharraf N. Google Gemini as a next generation AI educational tool: a review of emerging educational technology. Smart Learn Environ. 2024;11(1):22.
3. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023;29(8):1930-40. https://doi.org/10.1038/s41591-023-02448-8
4. GBD 2021 Stroke Risk Factor Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Neurol. 2024;23(10):973-1003. https://doi.org/10.1016/S1474-4422(24)00369-7
5. Lanas F, Seron P. Facing the stroke burden worldwide. Lancet Glob Health. 2021;9(3):e235-6. https://doi.org/10.1016/S2214-109X(20)30520-9
6. Rudberg AS, Berge E, Laska AC, Jutterström S, Näsman P, Sunnerhagen KS, et al. Stroke survivors' priorities for research related to life after stroke. Top Stroke Rehabil. 2021;28(2):153-8. https://doi.org/10.1080/10749357.2020.1789829
7. Chen L, Xiao LD, Chamberlain D, Newman P. Enablers and barriers in hospital-to-home transitional care for stroke survivors and caregivers: a systematic review. J Clin Nurs. 2021;30(19-20):2786-807. https://doi.org/10.1111/jocn.15807
8. Kazemi A, Azimian J, Mafi M, Allen KA, Motalebi SA. Caregiver burden and coping strategies in caregivers of older patients with stroke. BMC Psychol. 2021;9(1):51. https://doi.org/10.1186/s40359-021-00556-z
9. Tiwari S, Joshi A, Rai N, Satpathy P. Impact of stroke on quality of life of stroke survivors and their caregivers: a qualitative study from India. J Neurosci Rural Pract. 2021;12(4):680-8. https://doi.org/10.1055/s-0041-1735323
10. Naqvi IA, Montiel TC, Bittar Y, Hunter N, Okpala M, Johnson C, et al. Internet access and usage among stroke survivors and their informal caregivers: cross-sectional study. JMIR Form Res. 2021;5(3):e25123. https://doi.org/10.2196/25123
11. Maggio MG, Tartarisco G, Cardile D, Bonanno M, Bruschetta R, Pignolo L, et al. Exploring ChatGPT's potential in the clinical stream of neurorehabilitation. Front Artif Intell. 2024;7:1407905. https://doi.org/10.3389/frai.2024.1407905
12. Değerli Yİ, Özata Değerli MN. Using ChatGPT as a tool during occupational therapy intervention: a case report in mild cognitive impairment. Assist Technol. 2025;37(3):165-74. https://doi.org/10.1080/10400435.2024.2416495
13. Cheng HY. ChatGPT's attitude, knowledge, and clinical application in geriatrics practice and education: exploratory observational study. JMIR Form Res. 2025;9:e63494. https://doi.org/10.2196/63494
14. Kim S, Han DY, Bae J. Transforming Alzheimer's digital caregiving through large language models. Curr Alzheimer Res. 2024;21(7):503-16. https://doi.org/10.2174/0115672050301740241118044604
15. Fattah FH, Salih AM, Salih AM, Asaad SK, Ghafour AK, Bapir R, et al. Comparative analysis of ChatGPT and Gemini (Bard) in medical inquiry: a scoping review. Front Digit Health. 2025;7:1482712. https://doi.org/10.3389/fdgth.2025.1482712
16. Helbach J, Hoffmann F, Hecht N, Heesen C, Thomalla G, Wilfling D, et al. Information needs of people who have suffered a stroke or TIA and their preferred approaches of receiving health information: a scoping review. Eur Stroke J. 2025;10(1):5-21. https://doi.org/10.1177/23969873241272744
17. Muhrodji P, Wicaksono HDA, Satiti S, Trisnantoro L, Setyopranoto I, Vidyanti AN. Roles and problems of stroke caregivers: a qualitative study in Yogyakarta, Indonesia. F1000Res. 2021;10:380. https://doi.org/10.12688/f1000research.52135.2
18. Sidek NN, Kamalakannan S, Tengku Ismail TA, Musa KI, Ibrahim KA, Abdul Aziz Z, et al. Experiences and needs of the caregivers of stroke survivors in Malaysia-A phenomenological exploration. Front Neurol. 2022;13:996620. https://doi.org/10.3389/fneur.2022.996620
19. Mehdipanah R, Morgenstern LB, Kwicklis M, Smith MA, Case E, Lisabeth LD. Stroke caregiver needs in a population-based study. J Am Heart Assoc. 2025;14(19):e038434. https://doi.org/10.1161/JAHA.124.038434
20. Neo JRE, Ser JS, Tay SS. Use of large language model-based chatbots in managing the rehabilitation concerns and education needs of outpatient stroke survivors and caregivers. Front Digit Health. 2024;6:1395501. https://doi.org/10.3389/fdgth.2024.1395501
21. 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. https://doi.org/10.3389/fmed.2024.1477898
22. Mika AP, Martin JR, Engstrom SM, Polkowski GG, Wilson JM. Assessing ChatGPT responses to common patient questions regarding total hip arthroplasty. J Bone Joint Surg Am. 2023;105(19):1519-26. https://doi.org/10.2106/JBJS.23.00209
23. Sawamura S, Bito T, Ando T, Masuda K, Kameyama S, Ishida H. Evaluation of the accuracy of ChatGPT's responses to and references for clinical questions in physical therapy. J Phys Ther Sci. 2024;36(5):234-9. https://doi.org/10.1589/jpts.36.234
24. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155-63. https://doi.org/10.1016/j.jcm.2016.02.012

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251543

Tranexamic acid in arthroscopic repair of rotator cuff tears in the shoulder: a randomized trial

Aloísio Rosado Filho[1], Tércio Maia Sousa[1], João Nogueira Neto[1], Luciana Salles Branco-de-Almeida[1], Caio Marcio Barros de Oliveira[1], Marcelo Souza de Andrade[1], Ed Carlos Rey Moura[1], Almir Vieira Dibai Filho[2], Plínio da Cunha Leal[2,3]*

SUMMARY

OBJECTIVE: The aim of this study was to determine the effectiveness of tranexamic acid in improving intraoperative visual clarity in patients undergoing arthroscopic repair of rotator cuff tears.

METHODS: This prospective, randomized, double-blinded clinical trial was conducted in Brazil, between July 2020 and September 2022, including patients who underwent surgery for arthroscopic repair of rotator cuff tears.

RESULTS: In total, 20 patients were included in each group: control, oral, and intravenous. There were no statistically significant differences between groups regarding intraoperative interventions, including duration of surgery, anesthesia time, number of ablations, increase in irrigation pump pressure, or total volume of irrigating solution used. Likewise, no differences were observed between groups in the assessment of visual clarity, either subjective (initial, final, and median) or objective, although the groups that received tranexamic acid had slightly lower average turbidity compared to the control group.

CONCLUSION: Tranexamic acid use did not demonstrate benefits concerning the objective/subjective visual clarity between the evaluated groups.

KEYWORDS: Arthroscopy. Tranexamic acid. Visual clarity..

INTRODUCTION

Rotator cuff tears (RCT) are common in orthopedic practice[1], reaching 22% of patients over 65 years[2]. Surgery is a prevalent option in the treatment of arthroscopic rotator cuff repair, being the most commonly performed procedure[3].

Knowing how to control bleeding during arthroscopic procedures on the shoulder is vital; several studies have evaluated the reduction of blood loss and hematoma occurrence using tranexamic acid (TXA). TXA is a synthetic analogue of lysine that acts as an antifibrinolytic agent. It competitively blocks lysine receptors on plasminogen, thereby preventing fibrin clot degradation (fibrinolysis) and reducing bleeding[4].

Furthermore, the efficacy of TXA has been demonstrated in clinical trials, but their results still differ in visual clarity in shoulder arthroscopy[5].

Thus, this trial aimed to determine the effectiveness of TXA in improving intraoperative visual clarity in patients undergoing arthroscopic repair of RCT.

METHODS

Study design

This prospective, randomized, double-blind clinical trial followed CONSORT guidelines and was conducted between July 1, 2020, and September 30, 2022, in São Luís, Maranhão, Brazil.

Patients 18 years old or older, after 3 months of physiotherapy without pain improvement, who underwent arthroscopic repair of a rotator cuff tear rated between grade 0 and grade 2 in the Goutallier classification after magnetic resonance imaging were included.

Patients with a history of coagulopathy, altered coagulogram, renal or hepatic disorders, uncontrolled hypertension (systolic pressure >180 mmHg), allergic to anesthetic agents, or with irreparable/massive RCT with other surgical procedure needs, like tenotomy of the long head of the biceps tendon, were excluded.

[1]Federal University of Maranhão – São Luís (MA), Brazil.
[2]Federal University of Maranhão, Postgraduate Program in Physical Education – São Luís (MA), Brazil.
[3]Federal University of São Paulo, Postgraduate Program in Interdisciplinary Surgical Science – São Paulo (SP), Brazil.
*Corresponding author: plinio.cunha@ufma.br
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on July 28, 2025. Accepted on November 25, 2025.
Scientific Editor: Roseli Nomura

Data collection

Patient sociodemographic information, including gender, age, race/ethnicity, comorbidities, weight, and BMI, was systematically collected. During surgery, anesthesia and surgery duration, number of radiofrequency ablation uses, pump pressure increases, and irrigation fluid volume were recorded.

The measurement of subjective visual clarity was performed by analyzing the three-grade visual clarity scoring system performed during the surgical procedure every 15 min. The surgeon assigned grade 1 when visibility was low, with active bleeding, making it impossible to proceed with the surgery; grade 2 when visibility was reasonable, with moderate bleeding that interfered with vision but did not preclude the surgery; and grade 3 when the professional had good visibility, with no apparent blood presence[6].

Additionally, an objective visual clarity was quantified using computerized video analysis, in which surgical procedures were continuously recorded at 30 frames per second using a high-definition arthroscopic camera system. A custom image processing system was developed to generate histograms reflecting blood presence (calculated as the percentage of red pixels in the surgical field) and image sharpness (assessed through histogram analysis). This methodology provides a quantitative, observer-independent measure of visual clarity based on established principles of digital image processing. The images were independently analyzed by three surgeons to verify the correlation between the objective measures (% red pixels) and the subjective assessment of visual clarity.

Sample size

Sample size calculation was performed using G*Power software (v.3.1.9.4). The primary outcome was visual clarity assessed by the three-grade subjective scale. Based on a 4.3% incidence of significant intraoperative bleeding[7], an expected minimum clinically important difference of 0.5 points in the visual clarity scale between groups, a statistical power of 0.8 (80%), an alpha level of 0.05 (5%), and accounting for a 10% potential loss to follow-up, a minimum of 60 patients (20 per group) was required for adequate statistical power to detect clinically relevant differences.

Randomization

Randomization was conducted using a 1:1:1 ratio with software, assigning patients to control (CG), oral TXA (OG), or intravenous TXA (IG) groups, and concealment was ensured with sealed, numbered envelopes opened just before surgery. Each group had 20 patients, and participants and investigators collecting study data were blinded to group assignment until analysis.

Patients in the control group (CG) received ascorbic acid tablets, 250 mg each, at a dose of 20 mg/kg, 2 h before the procedure, as the placebo, and intravenous saline solution 10 min before the skin incision, with the same volume as if they were to receive intravenous TXA at a dose of 15 mg/kg. Patients in the oral group (OG) received TXA tablets, 250 mg each, at 20 mg/kg, 2 h before the procedure. To comply with the double-blind nature, patients in this group also received intravenous saline 10 min before skin incision, with the same volume as if they were to receive intravenous TXA at a dose of 15 mg/kg. In the intravenous group (IG), TXA was administered at 15 mg/kg 10 min before the skin incision. Participants in this group also received ascorbic acid (250 mg) orally at a 20 mg/kg dose 2 h before the procedure.

The oral dosage of 20 mg/kg was chosen to account for the approximately 45% oral bioavailability of TXA, while the intravenous dosage of 15 mg/kg follows standard clinical protocols[8]. Although these dosages are not strictly pharmacologically equivalent, the oral administration 2 h preoperatively allows achievement of therapeutic plasma concentrations (>10 μg/mL) at the time of incision, with the prolonged elimination half-life (approximately 11 h oral vs. 2 h IV) maintaining adequate levels throughout the procedure[9].

All drugs were administered by staff not involved in care or data collection. General anesthesia combined with an ultrasound-guided interscalene brachial plexus block was used for all patients. Anesthesia induction included IV midazolam, ropivacaine block, propofol, fentanyl, and rocuronium, maintained with sevoflurane. At the end of surgery, antiemetics and neuromuscular blockade reversal were administered, and anesthesia was discontinued when discharge criteria were met.

Surgical procedure

Patients were positioned in lateral decubitus, with the affected arm under longitudinal traction using a foam cuff and velcro support attached to the surgical table. The same surgeon performed all surgical procedures. The technique followed a double-row repair with medial and lateral anchors: subacromial decompression was performed first, followed by preparation of the rotator cuff footprint on the greater tubercle and suture fixation. Only 0.9% saline was used for irrigation, with fluid pressure set at 50 mmHg and increased by 20 mmHg increments if bleeding compromised visualization. Systolic and diastolic blood pressure were maintained within 100–130 and 50–80 mmHg, respectively, using vasopressors or hypotensive medications as needed. Extubation occurred once patients were awake and stable; after surgery, they remained under anesthetic recovery until meeting discharge criteria.

Outcomes

The primary outcome was visual clarity, which was assessed subjectively from the surgeon's perspective and objectively via computational image analysis.

Ethics statement

The study followed the guidelines of the Declaration of Helsinki, approved by the Ethics Committee of the Federal University of Maranhão (CAAE: 32371220.0.0000.5085) on December 17, 2020, and the data collection began after approval by the ethics committee. The study was registered in the Brazilian Registry of Clinical Trials (https://ensaiosclinicos.gov.br/rg/) (RBR-289z7vn) on February 2, 2022. The Free and Informed Consent Form, in writing, was obtained from all patients.

Statistical treatment

Data were analyzed using SPSS 21.0® (v.22) (EUA). The Shapiro-Wilk test was used to verify normality. The non-parametric Mann-Whitney and chi-square tests were applied to categorical variables to compare the differences in medians between groups. Differences were considered significant when the $p<0.05$.

RESULTS

In total, 60 patients completed the study: 20 per group (GC, GO, GI). No losses to follow-up occurred (Figure 1).

Female predominance was noted in TXA groups (55%). The median age was 59 years across groups. The GI group had the highest proportion of white participants (75%), and GC had more mixed-race participants (65%). Hypertension was the most common comorbidity (up to 50%).

No significant differences were observed in surgery duration (GC: 185 min [100–295]; GO: 162.5 min [80–250]; GI: 201.5 min [115–290]; p=0.292) or anesthesia time (GC: 255 min [175–415]; GO: 240 min [155–310]; GI: 270 min [170–365]; p=0.097). The number of radiofrequency ablations (p=0.932), pump pressure increases (p=0.419), and total irrigation volume (GC: 43 L [16–85]; GO: 30 L [12–83]; GI: 43 L [17–72]; p=0.119) also did not differ significantly (Table 1).

Subjective clarity scores at first assessment (p=0.415), last assessment (p=0.287), and median scores (GC: 2.8 [2.4–3.0]; GO: 2.8 [2.5–3.0]; GI: 2.9 [2.3–3.0]; p=0.985) showed no group differences. Objective image analysis yielded similar mean clarity values (GC: 1.945 [0.88–4.01]; GO: 1.680 [0.83–3.24]; GI: 1.775 [0.91–3.47]; p=0.421) (Table 2).

DISCUSSION

This randomized, double-blind clinical trial included 60 patients who received a placebo, oral, or intravenous TXA. The similarity in baseline characteristics between groups suggests that both agents can be used equivalently in arthroscopic rotator cuff repair.

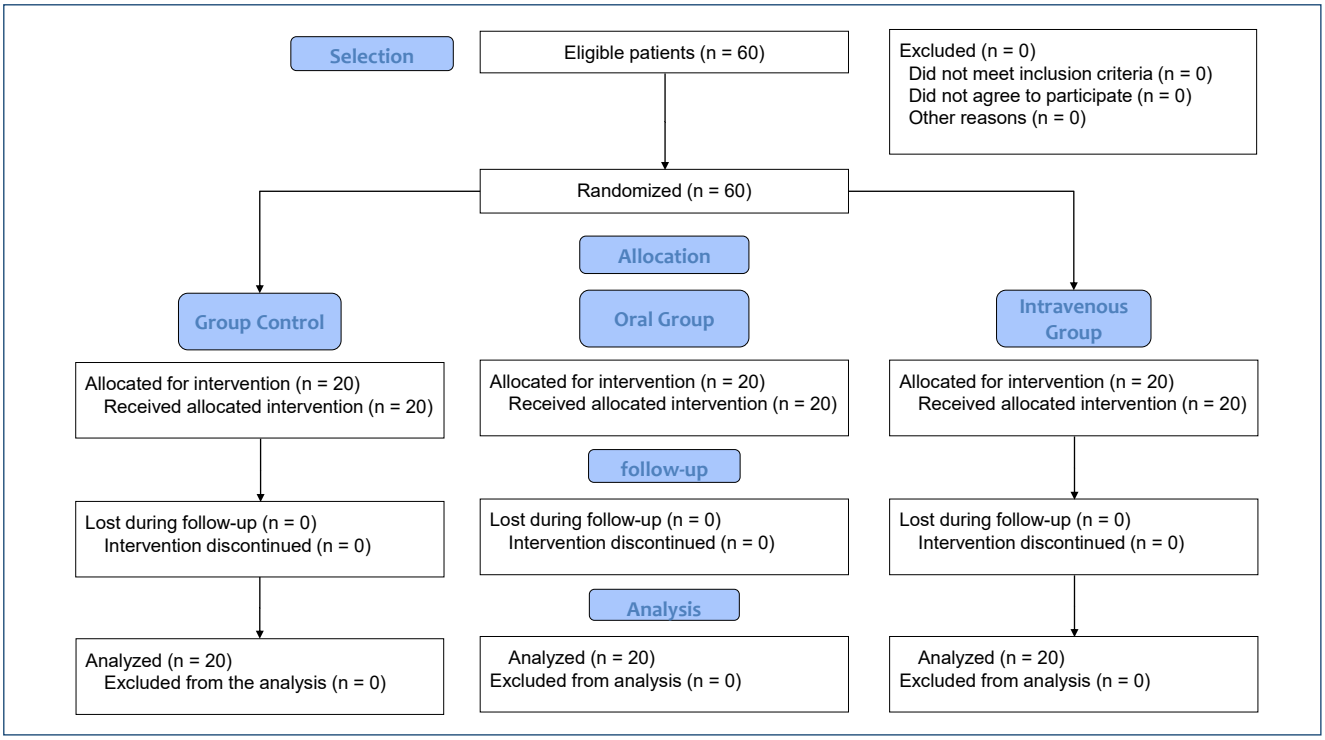


Figure 1. CONSORT flowchart.

Table 1. Sociodemographic, clinical history, anthropometry, and perioperative data of patients.

Variables	Groups			p-value§
	Control	Oral	Intravenous	
Sociodemographic, n (%)				
Gender				
Female	9 (45.0)	11 (55.0)	11 (55.0)	0.766
Male	11 (55.0)	9 (45.0)	9 (45.0)	
Age (years)				
39–49	4 (20.0)	4 (20.0)	5 (25.0)	0.571
50–59	7 (35.0)	10 (50.0)	5 (25.0)	
60 or more	9 (45.0)	6 (30.0)	10 (50.0)	
Median (Min–max)	59 (39–72)	56 (39–76)	59 (39–75)	0.878£
Race/ethnicity				
White	7 (35.0)	11 (55.0)	15 (75.0)	0.070
Black	0 (0.0)	1 (5.0)	0 (0.0)	
Brown	13 (65.0)	8 (40.0)	5 (25.0)	
Comorbidities				
SAH[a]	10 (50.0)	10 (50.0)	8 (40.0)	0.765
DM[b]	3 (15.0)	6 (30.0)	3 (15.0)	0.392
Dyslipidemia	2 (10.0)	2 (10.0)	6 (30.0)	0.147
Heart disease	2 (10.0)	3 (15.0)	4 (20.0)	0.676
Hypothyroidism	1 (5.0)	1 (5.0)	4 (20.0)	0.189
Anthropometry, med (min–max)				
Weight (Kg)	74 (61–98)	72.5 (59–98)	73.5 (53–98)	0.343
BMI[c] (Kg/m^2)	29.1 (21.4–37.4)	27.8 (23.6–33.2)	28.0 (16.9–35.6)	0.425
Perioperative data, med (min–max)				
Surgery duration (min)	185 (100–295)	162.5 (80–250)	201.5 (115–290)	0.292
Anesthesia duration (min)	255 (175–415)	240 (155–310)	270 (170–365)	0.097
Ablations	4 (0–11)	4.5 (0–8)	3 (0–9)	0.932
Increases in the pump	0 (0–6)	0 (0–6)	0 (0–7)	0.419
Irrigation fluid (L)	43 (16–85)	30 (12–83)	43 (17–72)	0.119

Systemic arterial hypertension; [b]diabetes mellitus; [c]body mass index; §chi-square; £ Kruskal-Wallis. SAH: systemic arterial hypertension; DM: diabetes mellitus; BMI: body mass index.

Table 2. Subjective and objective visual clarity in arthroscopic repair of rotator cuff tears.

Variables	Group			p-value
	Control	Oral	Intravenous	
Subjective visual clarity, n (%)				
First assessment				
Low	0 (0.0)	0 (0.0)	1 (5.0)	0.415§
Reasonable	2 (10.0)	3 (15.0)	5 (25.0)	
Good	18 (90.0)	17 (85.0)	14 (70.0)	
Last assessment				
Low	0 (0.0)	0 (0.0)	0 (0.0)	0.287§
Reasonable	2 (10.0)	6 (30.0)	4 (20.0)	
Good	18 (90.0)	14 (70.0)	16 (80.0)	
Subjective score med (min–max)	2.8 (2.4–3.0)	2.8 (2.5–3.0)	2.9 (2.3–3.0)	0.985£

Continue...

Table 2. Continuation.

Variables	Group			p-value
	Control	Oral	Intravenous	
Objective visual clarity, med (min–max)				
Average	1.945 (0.88–4.01)	1.680 (0.83–3.24)	1.775 (0.91–3.47)	0.421£
Standard deviation	2.135 (0.54–10.91)	2.075 (0.72–4.24)	2.550 (1.02–5.62)	0.599£
Median	2.240 (1.19–5.46)	2.485 (1.04–4.64)	2.605 (1.27–5.67)	0.559£

£Kruskal-Wallis; §chi-square.

Studies showing no benefit

Our findings align with several recent studies that found no significant benefit of TXA in improving visual clarity during arthroscopic rotator cuff repair. Nicholson et al.[10] found no difference in visualization between TXA and CGs (7.2±1.8 vs. 7.4±1.6, p=0.464) using surgeon-completed visual analog scales. Similarly, Bayram et al.[11] assessed visual clarity in arthroscopic rotator cuff repair and found no difference between epinephrine and TXA groups, with mean VAS scores of 7.6±1.62 versus 7.1±1.74 (p=0.59). Suter et al.[12] compared intravenous TXA, epinephrine in irrigation fluid, and their combination, finding mean visualization quality scores of 2.1 (±0.40), 2.1 (±0.52), 2.6 (±0.37), and 2.6 (±0.35) for control, TXA, epinephrine, and combination groups, respectively. Regression analysis showed TXA presence/absence did not affect visual clarity, suggesting TXA's hemostatic effect may be insufficient in moderate bleeding scenarios. Yoon et al.[13] evaluated three TXA administration methods (intravenous, topical, combined) in reverse shoulder arthroplasty and found no differences in hemoglobin drop, blood loss, drain output, or transfusion rates between groups.

Studies showing benefits

Conversely, several studies demonstrated significant benefits. Takahashi et al.[14] demonstrated significantly better visual clarity with 1 g TXA, showing a higher percentage of grade 3 visual clarity (75.6±11.2 vs. 68.1±13.4%, p=0.045). Bildik and Pehlivanoglu[15] investigated intra-articular TXA (250 mg/kg) and found significantly higher visual clarity (1.5±0.5 vs. 2.86±1.7; p<0.001), with a higher percentage of grade 1 clarity (78.1 vs. 32.2%) and lower grade 4 clarity (0 vs. 3.2%). Liu et al.[6] evaluated 1 g intravenous TXA and found significantly better visual clarity with higher grade 3 scores (53.7±18.9 vs. 40.5±22.1%, p=0.036) and better mean visual scores (2.5±0.2 vs. 2.3±0.3, p=0.048). Ersin et al.[16] found that visual clarity in the TXA group (10 mg/kg) with mean visual clarity scores=8.1/10 (range=7–10) was significantly better than in the control group (7/10 (range=5–9); p=0.018).

Analysis of discrepancies

These conflicting results can be explained by several methodological and contextual factors, like: (a) surgical stage specificity, in which Shin et al.[17] found that 1,000 mg intravenous TXA improved visual clarity during Stage I (synovectomy) compared to control (median grade 1 vs. 2, p=0.027) but showed no effect during acromioplasty, bursectomy, or greater tuberoplasty stages. This suggests TXA's impact is most pronounced during soft-tissue procedures with greater bleeding from the vascular synovial membrane; (b) dosing variability, as Wang et al.[18,19] demonstrated that combined intravenous plus topical TXA was significantly more effective than either alone, suggesting route and dose optimization may be critical; (c) studies demonstrating benefits may have included higher bleeding risk patients or more extensive procedures. For instance, Pauzenberger et al.[20] found that 1 g intravenous TXA reduced drain blood, total blood loss, pain, and hematoma formation in total shoulder arthroplasty, suggesting benefits may be more evident in procedures with greater surgical trauma. Also, the (d) assessment methodology, in which studies employed varying scales (3-point, 4-point, VAS 0–10), different assessment frequencies, and both objective and subjective measures. This heterogeneity in outcome measurement may partly explain conflicting results, and (e) procedure complexity and the type and extent of surgical intervention likely influence TXA efficacy. Procedures involving extensive synovectomy or more traumatic tissue manipulation may benefit more from antifibrinolytic therapy than routine rotator cuff repairs.

In summary, our study's absence of global differences in visual clarity aligns with the understanding that TXA benefits may be limited to specific high-bleeding scenarios or vascular procedures, rather than universally beneficial across all arthroscopic shoulder procedures. Standard rotator cuff repair, characterized by relatively controlled bleeding and less extensive synovial work, may not provide the clinical context where TXA's antifibrinolytic properties translate into measurable improvements in visual clarity.

Objective visual clarity

The use of computerized video analysis to objectively quantify visual clarity in this study represents an important methodological advancement in arthroscopic research. While subjective visual analog scales have been the standard for assessing intraoperative bleeding, they are inherently limited by observer variability and lack of standardization.

Recently, Birsel et al.[21] published a comprehensive validation study of an image processing system for objective quantification of intra-articular bleeding during arthroscopic rotator cuff repair. Their system, which uses color recognition algorithms combining HSV (hue, saturation, value) and CIELAB (International Commission on Illumination of L* for lightness, a* for the red–green axis, and b* for the yellow–blue axis) to calculate bleeding scores based on pixel ratios, demonstrated excellent agreement with experienced surgeon assessments (ICC 0.967, mean absolute error 0.56). Their validation study included 200 video clips evaluated by three senior surgeons, yielding high inter-observer (ICC0.96) and intra-observer (ICC 0.97) reliability. These validation results support the feasibility and reliability of objective visual clarity assessment methods similar to ours. Both approaches recognize that blood in arthroscopy acts like a red dye dissolving into arthroscopic fluid, requiring sophisticated color space analysis beyond simple RGB detection.

Limitations

This study has several important limitations that should be acknowledged. First, sample size and statistical power. Although our sample size calculation was based on established parameters and previous literature, the study may have been underpowered to detect small but potentially clinically relevant differences. Second, measurement bias. The subjective visual clarity assessment, although using a validated three-grade system recorded every 15 min, remains inherently subject to observer bias despite surgeon blinding to patient allocation. While the objective computerized image analysis partially mitigates this limitation, this specific system has not been previously validated in external published studies. Third, procedural variability. Despite standardized surgical technique, inherent variability in aspects such as the number of radiofrequency ablations and adjustments in irrigation pump pressure may independently influence visual clarity beyond the effect of TXA. Fourth, timing and dosing limitations. The study compared specific dosing regimens (20 mg/kg oral, 15 mg/kg IV) and timing protocols (oral 2 h preoperatively, IV 10 min before incision). Different doses, timing strategies, or combined administration routes might yield different results. Finally, procedure-specific considerations. Our findings apply specifically to standard rotator cuff repair in patients with Goutallier grade 0–2 fatty infiltration. Results may not generalize to more complex procedures or massive RCT, where bleeding patterns may differ substantially.

Despite these limitations, this study provides robust evidence from a well-designed, double-blind randomized controlled trial following CONSORT guidelines, with appropriate randomization, concealment, and blinding procedures. The findings contribute important information to clinical decision-making regarding TXA use in arthroscopic rotator cuff repair.

CONCLUSION

Current evidence suggests that while TXA is safe and well-tolerated, its routine use for visual clarity improvement in arthroscopic rotator cuff repair may not provide universal benefits. The decision to use TXA should be individualized based on patient bleeding risk, procedure complexity, and specific surgical requirements rather than applied universally to all arthroscopic rotator cuff repairs.

AUTHORS' CONTRIBUTIONS

ARF: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Writing – original draft. **TMS:** Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Writing – original draft. **JNN:** Conceptualization, Supervision, Validation, Visualization. **LSBA:** Conceptualization, Supervision, Validation, Visualization. **CMBO:** Conceptualization, Supervision, Validation, Visualization. **MSA:** Conceptualization, Supervision, Validation, Visualization. **ECRM:** Conceptualization, Supervision, Validation, Visualization. **AVDF:** Conceptualization, Supervision, Validation, Visualization. **PCL:** Data curation, Formal Analysis, Investigation, Methodology, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Moosmayer S, Tariq R, Stiris M, Smith HJ. The natural history of asymptomatic rotator cuff tears: a three-year follow-up of fifty cases. J Bone Joint Surg Am. 2013;95(14):1249-55. https://doi.org/10.2106/JBJS.L.00185
2. Fehringer EV, Sun J, VanOeveren LS, Keller BK, Matsen FA. Full-thickness rotator cuff tear prevalence and correlation with function and co-morbidities in patients sixty-five years and older. J Shoulder Elbow Surg. 2008;17(6):881-5. https://doi.org/10.1016/j.jse.2008.05.039
3. Warrender WJ, Brown OL, Abboud JA. Outcomes of arthroscopic rotator cuff repairs in obese patients. J Shoulder Elbow Surg. 2011;20(6):961-7. https://doi.org/10.1016/j.jse.2010.11.006
4. Camarasa MA, Ollé G, Serra-Prat M, Martín A, Sánchez M, Ricós P, et al. Efficacy of aminocaproic, tranexamic acids in the control of bleeding during total knee replacement: a randomized clinical trial. Br J Anaesth. 2006;96(5):576-82. https://doi.org/10.1093/bja/ael057
5. Goldstein K, Jones C, Kay J, Shin J, Sa D. Tranexamic acid administration in arthroscopic surgery is a safe adjunct to decrease postoperative pain and swelling: a systematic review and meta-analysis. Arthroscopy. 2022;38(4):1366-77.e9. https://doi.org/10.1016/j.arthro.2021.10.001
6. Liu YF, Hong CK, Hsu KL, Kuan FC, Chen Y, Yeh ML, et al. Intravenous administration of tranexamic acid significantly improved clarity of the visual field in arthroscopic shoulder surgery. A prospective, double-blind, and randomized controlled trial. Arthroscopy. 2020;36(3):640-7. https://doi.org/10.1016/j.arthro.2019.10.020
7. Vara AD, Koueiter DM, Pinkas DE, Gowda A, Wiater BP, Wiater JM. Intravenous tranexamic acid reduces total blood loss in reverse total shoulder arthroplasty: a prospective, double-blinded, randomized, controlled trial. J Shoulder Elbow Surg. 2017;26(8):1383-9. https://doi.org/10.1016/j.jse.2017.01.005
8. Chaunce y JM, Patel P. Tranexamic Acid. Em: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025. [cited on 2025 Oct 27]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK532909/
9. Kane Z, Picetti R, Wilby A, Standing JF, Grassin-Delyle S, Roberts I, et al. Physiologically based modelling of tranexamic acid pharmacokinetics following intravenous, intramuscular, sub-cutaneous and oral administration in healthy volunteers. Eur J Pharm Sci. 2021;164:105893. https://doi.org/10.1016/j.ejps.2021.105893
10. Nicholson TA, Kirsch JM, Churchill R, Lazarus MD, Abboud JA, Namdari S. The effect of tranexamic acid for visualization on pump pressure and visualization during arthroscopic rotator cuff repair: an anonymized, randomized controlled trial. J Shoulder Elbow Surg. 2022;31(11):2211-6. https://doi.org/10.1016/j.jse.2022.06.027
11. Bayram E, Yıldırım C, Ertürk AK, Yılmaz M, Atlıhan D. Comparison of the efficacy of irrigation with epinephrine or tranexamic acid on visual clarity during arthroscopic rotator cuff repair: a double-blind, randomized-controlled study. Jt Dis Relat Surg. 2021;32(1):115-21. https://doi.org/10.5606/ehc.2021.78393
12. Suter T, McRae S, Zhang Y, MacDonald PB, Woodmass JM, Mutter TC, et al. The effect of intravenous tranexamic acid on visual clarity in arthroscopic shoulder surgery compared to epinephrine and a placebo: a double-blinded, randomized controlled trial. J Shoulder Elbow Surg. 2024;33(3):628-39. https://doi.org/10.1016/j.jse.2023.10.001
13. Yoon JY, Park JH, Kim YS, Shin SJ, Yoo JC, Oh JH. Effect of tranexamic acid on blood loss after reverse total shoulder arthroplasty according to the administration method: a prospective, multicenter, randomized, controlled study. J Shoulder Elbow Surg. 2020;29(6):1087-95. https://doi.org/10.1016/j.jse.2020.02.013
14. Takahashi R, Kajita Y, Iwahori Y, Harada Y. Tranexamic acid administration for arthroscopic rotator cuff repair: a prospective, double-blind, randomized controlled trial. J Orthop Sci. 2023;28(2):328-32. https://doi.org/10.1016/j.jos.2021.11.015
15. Bildik C, Pehlivanoglu T. Arthroscopic rotator cuff repair performed with intra-articular tranexamic acid: could it provide improved visual clarity and less postoperative pain? A prospective, double-blind, randomized study of 63 patients. J Shoulder Elbow Surg. 2023;32(2):223-31. https://doi.org/10.1016/j.jse.2022.10.007
16. Ersin M, Demirel M, Büget Mİ, Edipoğlu İS, Atalar AC, Erşen A. The effect of intravenous tranexamic acid on visual clarity during arthroscopic rotator cuff repair: a randomized, double-blinded, placebo-controlled pilot study. Acta Orthop Traumatol Turc. 2020;54(6):572-6. https://doi.org/10.5152/j.aott.2020.19164
17. Shin HJ, You HS, Lee K, Je LG, Kim SG, Yang KS, et al. Intravenous tranexamic acid improves visual clarity during synovectomy in patients undergoing arthroscopic rotator cuff repair: a double-blind, randomized controlled study. Arthroscopy. 2024;40(5):1409-19. https://doi.org/10.1016/j.arthro.2023.10.019
18. Wang TC, Guo JL, Tian QP, Deng HP, Yin B, Xiao Z, et al. Application of tranexamic acid in shoulder arthroscopic surgery: a randomised controlled trial. Chin Med Sci J. 2023;38(4):273-8. https://doi.org/10.24920/004295
19. Wang H, Ouyang Y, Chen Y, Zeng L. Clinical application of tranexamic acid in arthroscopic rotator cuff repair surgery: a randomized controlled trial. Medicine (Baltimore). 2024;103(26):e38515. https://doi.org/10.1097/MD.0000000000038515
20. Pauzenberger L, Domej MA, Heuberer PR, Hexel M, Grieb A, Laky B, et al. The effect of intravenous tranexamic acid on blood loss and early post-operative pain in total shoulder arthroplasty. Bone Joint J. 2017;99-B(8):1073-9. https://doi.org/10.1302/0301-620X.99B8.BJJ-2016-1205.R1
21. Birsel O, Zengin U, Eren I, Ersen A, Semiz B, Demirhan M. Validation of novel image processing method for objective quantification of intra-articular bleeding during arthroscopic procedures. J Imaging. 2025;11(2):40. https://doi.org/10.3390/jimaging11020040

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251578

The effect of 30° and 70° arthroscope use on femoral and tibial tunnel placement in arthroscopic anterior cruciate ligament reconstruction surgery

Öner Kilinc[1]*, Fatih Günaydin[1], İdris Demirtas[1], Bülent Sakarya[1]

SUMMARY

OBJECTIVE: The success of anterior cruciate ligament reconstruction depends on anatomical tunnel placement. The aim of this was to compare the effects of 30° and 70° arthroscopes on femoral and tibial tunnel positioning and their relationship with clinical outcomes.

METHODS: This retrospective study analyzed 55 patients who underwent anterior cruciate ligament reconstruction between 2020 and 2024 (30° arthroscope, n=28; 70° arthroscope, n=27). All procedures were performed by the same surgeon using hamstring tendon grafts with suspensory femoral fixation and bioabsorbable tibial screws. Femoral and tibial tunnel positions were evaluated on postoperative computed tomography using the Bernard-Hertel and Staubli-Rauschning methods, respectively, with measurements performed by four orthopedic surgeons; clinical outcomes were assessed at ≥6 months using the International Knee Documentation Committee score.

RESULTS: Femoral tunnels were significantly more posterior and distal in the 70° group (depth 29.4 vs. 38.7%; height 33.7 vs. 43.1%; both p<0.001). Tibial tunnel placement showed no significant difference (p=0.49). International Knee Documentation Committee scores were higher in the 70° group (89.2±5.4 vs. 84.6±6.8; p=0.02). Correlation and regression analyses confirmed femoral tunnel position as an independent predictor of functional outcome. Receiver operating characteristic curve analysis showed area under the curve values of 0.75 (95%CI 0.69–0.79) for depth and 0.71 (95%CI 0.67–0.78) for height.

CONCLUSION: Use of a 70° arthroscope in anterior cruciate ligament reconstruction enables more anatomical femoral tunnel placement and improves functional outcomes, while tibial tunnel positioning is unaffected. Arthroscopic viewing angle should be regarded not only as a technical factor but also as a determinant of clinical success.

KEYWORDS: Anterior cruciate ligament reconstruction. Arthroscopy. Anterior cruciate ligament injuries. Arthroscopic surgery.

INTRODUCTION

Anterior cruciate ligament (ACL) injuries are among the most common orthopedic problems that compromise the anterior and rotational stability of the knee. In active individuals, these injuries can lead to significant functional loss, and arthroscopic anatomical reconstruction surgery is considered the gold standard treatment[1].

The anatomical placement of the graft is critical for surgical success[2,3]. During surgery, femoral and tibial tunnels are created using anatomical landmarks, guiding instruments, or fluoroscopy. The positioning of these tunnels plays a decisive role in both the biomechanical effectiveness of the reconstruction and the clinical outcomes[3,4].

For assessing the anatomical position of the femoral tunnel, the "quadrant" method described by Bernard and Hertel is widely used[5]. The accuracy of tunnel placement largely depends on the arthroscopic viewing angle. While traditionally used 30° arthroscopes may be insufficient for visualizing certain anatomical areas, 70° arthroscopes provide clearer visualization of the posterior border of the femoral footprint, enabling surgeons to achieve a more anatomical tunnel placement[6,7].

Tibial tunnel placement, on the other hand, is generally guided by fixed anatomical landmarks and is therefore less affected by the arthroscopic viewing angle[8]. Nonetheless, the anteroposterior positioning of the tibial tunnel is also crucial for the functional performance of the graft[9].

This study aims to radiologically compare the effects of 30° and 70° arthroscopes on femoral and tibial tunnel placement in arthroscopic ACL reconstruction surgery and to evaluate their relationship with clinical scores.

[1]Mersin City Education and Research Hospital, Department of Orthopaedics and Traumatology – Mersin, Turkey.
*Corresponding author: dronerkilinc@hotmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on September 05, 2025. Accepted on January 12, 2026.
Scientific Editor: José Maria Soares Júnior

METHODS

In this retrospective comparative study, 55 patients who underwent arthroscopic ACL reconstruction between 2020 and 2024 were evaluated. All operations were performed by the same surgeon using ipsilateral hamstring tendon grafts and the manual anteromedial portal technique. The same anteromedial portal was used for both the 30° and 70° arthroscopes without any modification in portal placement; the 70° scope was employed solely to enhance visualization. The use of the 30° versus 70° arthroscope was not randomized; the preference for the 70° arthroscope reflected the surgeon's technical choice to obtain a wider and clearer visualization of the femoral footprint, rather than any patient-related characteristics. Suspensory fixation devices were used for femoral tunnel fixation, and bioabsorbable screws were used for tibial tunnel fixation. Patients were divided into two groups based on the type of arthroscope used during surgery: Group A (30° arthroscope, n=28) and Group B (70° arthroscope, n=27).

Patients with a primary ACL tear, who had undergone postoperative knee computed tomography (CT) scans and had a minimum of 6 months of follow-up, along with complete clinical and demographic data, were included in the study. Exclusion criteria were revision surgery, concomitant ligament injuries, and inadequate CT imaging.

Femoral tunnel position was evaluated using the Bernard-Hertel quadrant method, while tibial tunnel position was analyzed using the Staubli-Rauschning method. The measurement procedures and the tunnel configurations of the groups are shown in Figure 1. Radiographic measurements were independently performed by four orthopedic surgeons. Any discrepancies among observers were resolved through a consensus-based approach, and the final value used in the analysis represented the shared agreement of all observers. At least 6 months postoperatively, all patients were evaluated with the International Knee Documentation Committee (IKDC) form, and the scores were analyzed.

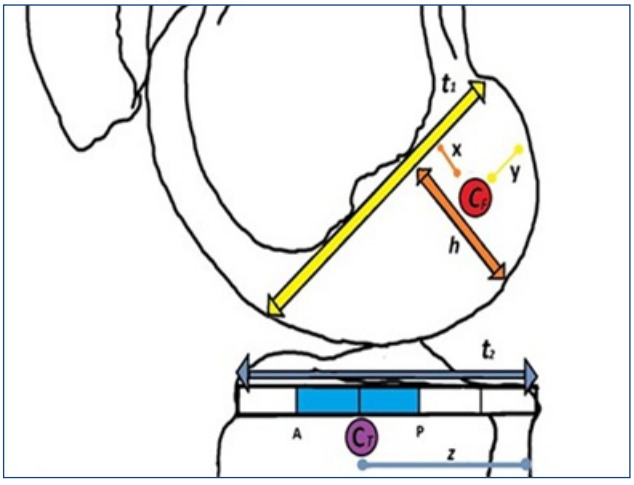


Figure 1. Diagram demonstrating the method used to measure the femoral and tibial tunnels. Evaluation of the femoral tunnel using the Bernard-Hertel method and the tibial tunnel using the Staubli-Rauschning method. Cf: Femoral tunnel center; x: distance to the Blumensaat line; y: distance to the posterior border; t1: length of the Blumensaat line; h: distance between the Blumensaat line and the distal border; Ct: Tibial tunnel center; A: anterior border of the ACL; P: posterior border of the ACL; t2: sagittal length of the tibia; z: distance from the tibial tunnel center to the posterior border; x/h: femoral height; y/t1: femoral depth; z/t2: tibial depth.

The study was approved by the Mersin University Clinical Research Ethics Committee (Approval No: 2023/730, dated 01/11/2023), and informed consent was obtained from all patients. Based on power analysis performed with G*Power 3.1 software (α=0.05, 1–β=0.80, effect size d=0.8), a minimum of 26 patients per group was calculated to be required.

Statistical analysis

Statistical analyses were conducted using SPSS v25.0 software. Data distribution was assessed with the Kolmogorov-Smirnov test. An independent-samples t-test was used for parametric variables, a chi-square test for categorical variables, and a $p<0.05$ was considered statistically significant.

RESULTS

The 55 patients included in the study were divided into two groups according to the type of arthroscope used: Group A (30° arthroscope, n=28) and Group B (70° arthroscope, n=27). The groups were comparable in terms of age, sex, affected side, dominant side, and follow-up duration. The mean age was 27.4±5.8 years in Group A and 28.1±6.3 years in Group B, with no statistically significant difference (p=0.62). The proportion of male patients was 75% in Group A and 71% in Group B (p=0.77). The minimum follow-up period was 6 months for all patients, with mean follow-up durations of 9.1±2.3 months in Group A and 10.2±3.1 months in Group B. The dominant side was the right in 23 patients (82.1%) in Group A and 21 patients (77.8%) in Group B (p=0.69); the number of patients with the right knee affected was 17 (61%) and 18 (66%), respectively (p=0.73). The demographic, surgical, and clinical characteristics of the patients are summarized in Table 1.

When femoral tunnel positions were evaluated using the Bernard-Hertel quadrant method, the mean depth value of the femoral tunnel center was 29.4%±3.9 and the mean height value was 33.7%±4.2 in Group B. In Group A, these values were 38.7%±5.1 and 43.1%±4.8, respectively. Statistically significant differences were observed between the two groups in both depth and height parameters ($p<0.001$ for both). A shallower (i.e., more anatomical) femoral tunnel placement was positively correlated with higher IKDC scores. The locations of tunnel placements for the groups are demonstrated in Figure 2.

Table 1. Comparison of demographic, surgical, and clinical characteristics between the 30° and 70° arthroscope groups.

Parameter	Group A (30°)	Group B (70°)	p-value
Age, years (mean±SD)	27.4±5.8	28.1±6.3	0.62
Male gender (n, %)	21 (75)	19 (71)	0.77
Right knee involvement (n, %)	17 (61)	18 (66)	0.73
Left knee involvement (n, %)	11 (39)	9 (34)	0.73
Dominant side (n, %)	23 (82.1)	21 (77.78)	0.69
Follow-up duration, months (mean±SD)	9.1±2.3	10.2±3.1	0.77
Femoral tunnel depth (mean±SD)	38.7±5.1	29.4±3.9	<0.001*
Femoral tunnel height (mean±SD)	43.1±4.8	33.7±4.2	<0.001*
Tibial tunnel position (mean±SD)	42.9±4.3	43.3±3.8	0.49
IKDC score (mean±SD)	84.6±6.8	89.2±5.4	0.02*

IKDC: International Knee Documentation Committee; SD: standard deviation. *p<0.05.

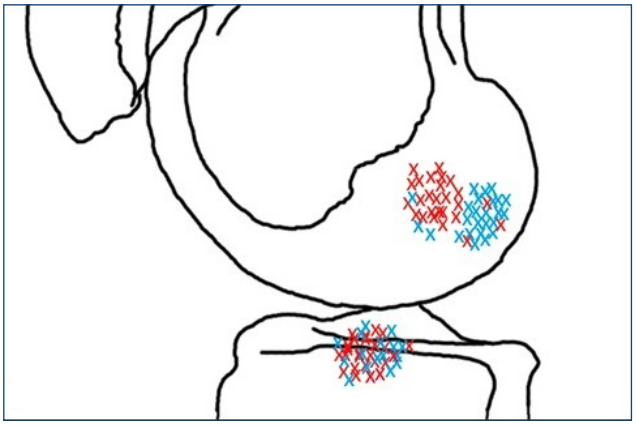


Figure 2. Illustration showing the tunnel configurations for the femur and tibia in the groups. Red crosses represent Group A (30° arthroscope), whereas blue crosses represent Group B (70° arthroscope).

Tibial tunnel placement, evaluated using the Staubli-Rauschning method, revealed an anteroposterior position ratio of 42.9%±4.3 in Group A and 43.3%±3.8 in Group B, with no significant difference between groups (p=0.49). Postoperative IKDC scores were 84.6±6.8 in Group A and 89.2±5.4 in Group B, and this difference was statistically significant (p=0.02).

Femoral tunnel depth demonstrated an AUC of 0.75 (95%CI 0.69–0.79), and femoral tunnel height demonstrated an AUC of 0.71 (95%CI 0.67–0.78); both values were observed as predictors of clinical outcomes.

DISCUSSION

This study aimed to evaluate the effects of 30° and 70° arthroscopes on femoral and tibial tunnel placement in arthroscopic ACL reconstruction surgery and to investigate the relationship of these placements with clinical outcomes. The findings demonstrated that the use of a 70° arthroscope provides significant advantages, particularly with respect to femoral tunnel positioning. In the 70° arthroscope group, the femoral tunnel was located significantly more posterior and distal, which corresponds to a more anatomical positioning. This difference can be attributed to the ability of the wide-angle arthroscope to provide clearer visualization of the posterior border of the femoral footprint. In the literature, Bedi et al. reported that the 70° arthroscope is superior to the conventional 30° arthroscope in visualizing anatomical structures behind the intercondylar notch[6]. Similarly, pioneering studies by Gillquist and Schneider emphasized the direct impact of portal placement on the quality of arthroscopic visualization[7,8].

From a clinical perspective, this anatomical advantage was reflected in patient outcomes. IKDC scores were significantly higher in patients operated on with a 70° arthroscope, indicating that more anatomical femoral tunnel placement supports better postoperative functional recovery. Correlation analyses further confirmed this relationship. Significant positive correlations were found between femoral tunnel depth and height parameters and IKDC scores. Regression analyses revealed that femoral tunnel placement was an independent predictor. These findings are consistent with Zhang et al., who emphasized the association between femoral tunnel position, knee stability, and functional outcomes[3].

Our study demonstrated that femoral tunnel parameters showed an association with IKDC outcomes. Receiver operating characteristic (ROC) analysis indicated that femoral tunnel depth (area under the curve [AUC]=0.75, 95%CI 0.69–0.79) and height (AUC=0.71, 95%CI 0.67–0.78) had only moderate discriminative capacity. These findings suggest that a more anatomical femoral tunnel placement may contribute to early functional improvement, although the predictive value of these parameters remains limited and should be interpreted with caution.

On the other hand, no significant difference was observed between groups in terms of tibial tunnel placement.

This can be explained by the fact that tibial tunnels are guided by more fixed and predictable anatomical landmarks and that the arthroscopic viewing angle has limited influence on the posterior tibial area. Measurements obtained using the percentage method described by Staubli and Rauschning yielded similar results in both groups[10]. Nevertheless, opposing views in the literature suggest that tibial tunnel placement is not constant. Iriuchishima and Goto demonstrated that the position of the tibial spine is a determinant of tunnel orientation, while Pedneault and colleagues reported that, in patients undergoing anatomic reconstruction, deviations may occur due to these anatomical variations[11,12]. In addition, Montreuil et al. suggested the use of biplanar stereoradiography and grid systems for more precise evaluation of tibial tunnel placement[13].

It should also be noted that tibial tunnel drilling can affect not only graft positioning but also meniscal integrity. In a cadaveric study, LaPrade et al. demonstrated that placing the tunnel over the anterior root of the lateral meniscus may compromise its structural integrity[14]. Therefore, tibial tunnel placement should be carefully considered not only in terms of graft trajectory but also in relation to surrounding structures. In this context, the 70° arthroscope was also observed to be more effective in visualizing the anterior margins of the menisci.

In our study, ipsilateral hamstring tendon grafts were used in all patients, with suspensory fixation on the femoral side and bioabsorbable screws on the tibial side. The studies of Hatipoğlu et al. and Alomari et al. have shown that this graft type yields reliable and successful clinical outcomes[15,16]. Previous studies also reported that bioabsorbable screws provide similar clinical success rates to metallic screws and demonstrate high biocompatibility[17]. Samuelsson et al., in a systematic review, emphasized the importance of maintaining homogeneity in graft and fixation methods when comparing surgical techniques[18]. Accordingly, the use of a uniform graft and fixation method in our study allowed for an isolated evaluation of the effect of arthroscopic viewing angle on femoral tunnel placement, thereby enhancing the reliability of the group comparison and strengthening the relationship between findings and arthroscope type.

Although IKDC scores were statistically higher in the 70° arthroscope group, the magnitude of this early difference may have limited clinical impact. However, the more anatomical femoral tunnel positioning obtained with the 70° arthroscope may contribute to functional improvement, which could become more evident with longer-term follow-up.

This study has certain limitations. First, the retrospective design and limited sample size restrict the generalizability of the findings. Although the power analysis confirmed that the sample size was adequate to detect major differences in femoral tunnel parameters, the overall number of patients was still limited. Therefore, smaller or more subtle differences — particularly in secondary outcomes such as tibial tunnel positioning — may not have been fully captured. This may restrict the generalizability of the results. Due to the retrospective design, adjustments for variables such as preoperative activity level, concomitant meniscal pathology, or rehabilitation protocols were not possible. However, the demographic and surgical similarities between groups likely minimized their potential influence. All procedures were performed by a single surgeon with extensive arthroscopic experience, and the use of the 70° arthroscope reflected a technical preference rather than a learning-curve effect. The mean follow-up period of 9–10 months reflects early postoperative outcomes and does not allow assessment of long-term functional results or graft integrity. Longer-term follow-up studies are needed to better understand these effects.

The exclusive use of hamstring autografts and uniform fixation methods minimizes surgical variability but limits the generalizability of the findings to other graft choices such as bone–patellar, tendon–bone, or allografts. Graft healing and biological integration could not be evaluated with advanced imaging modalities such as MRI. As suggested by Bird and Steiner, three-dimensional CT and high-resolution imaging techniques could provide a more precise analysis of tunnel placement and graft integrity[19,20]. In addition, the use of a 70° arthroscope may present technical challenges for surgeons, including a learning curve, portal positioning, and operative time. Therefore, future prospective, randomized, multicenter studies that incorporate variables such as surgeon experience, operative time, graft type, and imaging modalities are warranted to enhance the reliability and clinical applicability of the findings.

CONCLUSION

This study demonstrated that the use of a 70° arthroscope in arthroscopic ACL reconstruction allows for more anatomical femoral tunnel placement, which in turn translates into improved clinical outcomes. The improvement in femoral tunnel positioning was associated with higher IKDC scores. In contrast, tibial tunnel placement was not influenced by arthroscope type. These findings indicate that the arthroscopic viewing angle is an important factor affecting not only surgical technique but also patient outcomes. The enhanced visualization provided by the 70° arthroscope may contribute to more anatomical tunnel placement.

Future studies with larger samples, prospective designs, and comprehensive analyses—including surgeon experience and long-term results—will be beneficial in validating these findings.

AUTHORS' CONTRIBUTIONS

ÖK: Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **FG:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **İD:** Writing – review & editing. **BS:** Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Manderle BJ, Beletsky A, Gorodischer T, Chahla J, Cancienne JM, Vadhera AS, et al. Transtibial anterior cruciate ligament reconstruction: tips for a successful anatomic reconstruction. Arthrosc Tech. 2021;10(12):e2783-8. https://doi.org/10.1016/j.eats.2021.08.025
2. Kurihara S, Yanagisawa S, Takahashi T, Hagiwara K, Hatayama K, Takase R, et al. Increased bone plug depth from the joint increases tunnel enlargement in anterior cruciate ligament reconstruction using bone-patellar tendon-bone autograft with suspensory femoral fixation. Arthrosc Sports Med Rehabil. 2023;5(4):100755. https://doi.org/10.1016/j.asmr.2023.100755
3. Zhang Z, Meng B, Li W, Cao J. The role of navigation technology in anterior cruciate ligament reconstruction bone tunnel positioning. J Robot Surg. 2025;19(1):90. https://doi.org/10.1007/s11701-025-02254-z
4. Chen T, Chen J, Li X, He Y, Peng Q, Chen H. A comparative study on 3D printing-assisted arthroscopic IDEAL point femoral tunnel positioning for anterior cruciate ligament reconstruction versus conventional arthroscopic positioning. BMC Musculoskelet Disord. 2024;25(1):481. https://doi.org/10.1186/s12891-024-07591-y
5. Bernard M, Hertel P, Hornung H, Cierpinski T. Femoral insertion of the ACL. radiographic quadrant method. Am J Knee Surg. 1997;10(1):14-21. PMID: 9051173
6. Bedi A, Dines J, Dines DM, Kelly BT, O'Brien SJ, Altchek DW, et al. Use of the 70° arthroscope for improved visualization with common arthroscopic procedures. Arthroscopy. 2010;26(12):1684-96. https://doi.org/10.1016/j.arthro.2010.04.070
7. Tanpowpong T, Itthipanichpong T, Limskul D, Tantimethanon T, Jenvorapoj S. How to decrease suture tangling when performing supraspinatus repair with a single-working portal technique. Arthrosc Tech. 2022;11(7):e1219-22. https://doi.org/10.1016/j.eats.2022.03.008
8. Vijayan S, Kyalakond H, Kulkarni MS, Aroor MN, Shetty S, Bhat V, et al. Clinical outcome of anterior cruciate ligament reconstruction with modified transtibial and anteromedial portal. Musculoskelet Surg. 2023;107(1):37-45. https://doi.org/10.1007/s12306-021-00727-6
9. Moon DK, Jo HS, Lee DY, Kang DG, Won HC, Seo MS, et al. Anterior cruciate ligament femoral-tunnel drilling through an anteromedial portal: 3-dimensional plane drilling angle affects tunnel length relative to notchplasty. Knee Surg Relat Res. 2021;33(1):13. https://doi.org/10.1186/s43019-021-00092-5
10. Stäubli HU, Rauschning W. Tibial attachment area of the anterior cruciate ligament in the extended knee position. Anatomy and cryosections in vitro complemented by magnetic resonance arthrography in vivo. Knee Surg Sports Traumatol Arthrosc. 1994;2(3):138-46. https://doi.org/10.1007/BF01467915
11. Iriuchishima T, Goto B. Tibial spine location influences tibial tunnel placement in anatomical single-bundle anterior cruciate ligament reconstruction. J Knee Surg. 2022;35(3):294-8. https://doi.org/10.1055/s-0040-1713736
12. Pedneault C, Laverdière C, Hart A, Boily M, Burman M, Martineau PA. Evaluating the accuracy of tibial tunnel placement after anatomic single-bundle anterior cruciate ligament reconstruction. Am J Sports Med. 2019;47(13):3187-94. https://doi.org/10.1177/0363546519873633
13. Montreuil J, Saleh J, Cresson T, Guise JA, Lavoie F. Tibial tunnel placement in ACL reconstruction using a novel grid and biplanar stereoradiographic imaging. Orthop J Sports Med. 2021;9(3):2325967121989369. https://doi.org/10.1177/2325967121989369
14. LaPrade CM, Smith SD, Rasmussen MT, Hamming MG, Wijdicks CA, Engebretsen L, et al. Consequences of tibial tunnel reaming on the meniscal roots during cruciate ligament reconstruction in a cadaveric model, Part 1: the anterior cruciate ligament. Am J Sports Med. 2015;43(1):200-6. https://doi.org/10.1177/0363546514554769
15. Hatipoğlu MY, Bircan R, Özer H, Selek HY, Harput G, Baltacı YG. Radiographic assessment of bone tunnels after anterior cruciate ligament reconstruction: a comparison of hamstring tendon and bone-patellar tendon-bone autografting technique. Jt Dis Relat Surg. 2021;32(1):122-8. https://doi.org/10.5606/ehc.2021.75694
16. Alomari MS, Ghaddaf AA, Abdulhamid AS, Alshehri MS, Ashraf M, Alharbi HH. Single bundle versus double bundle anterior cruciate ligament reconstruction: a systematic review and meta-analysis. Indian J Orthop. 2022;56(10):1669-84. https://doi.org/10.1007/s43465-022-00718-0
17. Xu B, Yin Y, Zhu Y, Yin Y, Fu W. Clinical and radiographic outcomes of bioabsorbable screw fixation in ACL reconstruction: comparison with metal screws. *Knee Surg Relat Res*. 2013;25(3):112-7.
18. Samuelsson K, Andersson D, Ahldén M, Fu FH, Musahl V, Karlsson J. Trends in ACL reconstruction surgery: a systematic review. *Knee Surg Sports Traumatol Arthrosc*. 2010;18(1):19-34.
19. Bird JH, Carmont MR, Dhillon M, Smith N, Brown C, Thompson P, et al. Validation of a new technique to determine midbundle femoral tunnel position in anterior cruciate ligament reconstruction using 3-dimensional computed tomography analysis. Arthroscopy. 2011;27(9):1259-67. https://doi.org/10.1016/j.arthro.2011.03.077
20. Steiner M. Anatomic single-bundle ACL reconstruction. Sports Med Arthrosc Rev. 2009;17(4):247-51. https://doi.org/10.1097/JSA.0b013e3181c0ccf8

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251605

Immersive learning in obstetrics: a 360-degree video for teaching the management of premature rupture of membranes

Eduardo Moreira Nader[1], Ana Beatriz Ramos Antiqueira[1], Roseli Mieko Yamamoto Nomura[1]*

SUMMARY

OBJECTIVE: The aim of this study was to evaluate the performance and perception of a 360-degree video as an educational tool for medical students in the management of premature rupture of membranes.

METHODS: A prospective interventional study was conducted with fifth-year medical students. Participants watched a 10-min, 360-degree video simulating a complete premature rupture of membranes case, from anamnesis to prescription. Student perceptions were assessed using a questionnaire with a five-point Likert scale, covering technical quality, comprehension of clinical procedures, and pedagogical impact. A subgroup analysis compared outcomes based on the viewing device (mobile phone vs. computer).

RESULTS: A total of 60 medical students were included in this study. The video intervention received a highly favorable evaluation. Students reported high rates of comprehension for key clinical decision-making elements, including the indication for hospital admission (60 students, 100%) and the rationale for antibiotic therapy (59 students, 98.3%). The experience positively influenced student motivation (53 students, 88.3%) and heightened their interest in obstetrics (45 students, 75.0%). No statistically significant ($p>0.05$) differences were found between the proportion of positive responses (agree and strongly agree) according to the device used to watch the 360-degree video: mobile phone (n=34, 56.7%) vs. computer (n=26, 43.3%).

CONCLUSION: The use of 360-degree video is an effective and well-received pedagogical tool for teaching complex obstetric scenarios. Its comparable effectiveness on both mobile and computer platforms underscores its potential as an accessible and scalable resource to better prepare students for the practical challenges of the obstetrics and gynecology clerkship.

KEYWORDS: Education, medical. Simulation training. Virtual reality. Obstetrics. Fetal membranes, premature rupture.

INTRODUCTION

Since the beginning of the new millennium, educators around the world have been exploring new teaching methods, recognizing that students born in this century—commonly referred to as digital natives—require more interactive and diverse pedagogical approaches[1]. This challenge is particularly pronounced in medical education, where the ability to capture and maintain students' attention has become crucial to an effective teaching–learning process[2].

In this context, the use of innovative technologies, such as 360-degree video, has emerged as a promising tool for improving students' visual and auditory immersion in the learning environment[3]. By allowing simultaneous viewing of all perspectives, 360-degree video transcends the limitations of traditional two-dimensional videos, aiming for greater interactivity and viewer engagement[4].

The importance of new technologies in medical education has become even more relevant after the COVID-19 pandemic, which accelerated the transition from face-to-face teaching models to distance learning models, highlighting the fundamental role of digital media in medical training[5]. In the specific field of obstetrics, where students' exposure to practical situations can have implications for maternal and fetal health, virtual simulation has been recognized as a beneficial and safe approach[6,7].

The visual and auditory immersion provided by 360-degree video presents itself as an alternative for teaching, capable of overcoming distractions, and as a valuable tool for engaging students[8]. Recent studies indicate that the use of 360-degree videos in health education promotes improvements in student motivation and self-confidence. These improvements are attributed to a greater sense of "presence" and immersion in the learning environment, which are fundamental to the consolidation of skills[9,10].

Although there is growing adoption of this technology for emergency training[10], there is a gap in teaching the management of obstetric conditions such as premature rupture of

[1]Universidade Federal de São Paulo, Escola Paulista de Medicina, Department of Obstetrics, Division of Experimental and Physiologic Obstetrics – São Paulo (SP), Brazil.

*Corresponding author: roseli.nomura@unifesp.br

Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.

Received on October 02, 2025. Accepted on February 12, 2026.

Scientific Editor: José Maria Soares Júnior

membranes (PROM). Therefore, this study focuses on the use of 360-degree video as a teaching resource, exploring its performance as a facilitator of obstetric care learning. The objective is to evaluate the performance and perception of a 360-degree video as an educational tool for medical students in the treatment of PROM.

METHODS

This is a prospective interventional study addressing the simulation of management of PROM. The research protocol was approved by the Institutional Research Ethics Committee (CAAE Number 57376822.0.0000.5505), and all the participants signed a consent form.

A 10-min 360-degree video was produced using a Samsung Gear 360 camera. The video was developed at the Simulation Center at the Federal University of São Paulo, in a facility equipped for virtual media creation. In the simulation session, a NoelleS575, a full-sized female anthropomorphic birth simulator, and a model of cervical dilation (Life/form® Cervical Dilatation and Effacement Simulator) were used. The room in the simulation lab is specially prepared to manage obstetric cases. For the scenarios that simulate a woman's admission, standardized speeches were prepared with specific topics challenging the student caring for the simulated patient complaining of rupture of membranes. This scenario was developed with the participation of an actress (a graduate student) representing a woman at 39 weeks' gestation who tests positive for Group B streptococcus. She complains of fluid loss from her vagina that started 8 h before and is now experiencing contractions every 30 min. The doctor conducts an anamnesis to establish the full extent of the situation, including the woman's expectations. A physical examination was performed on the full-sized female simulator to verify maternal and fetal status and to reach a clinical diagnosis by conducting a speculum examination and using diagnostic tests. Furthermore, the Bishop's index was calculated to determine obstetric management and perform a medical prescription. In this scenario, a 2 cm cervical dilation effacement simulator and a medical prescription form were used. In addition, the interpretation of the cardiotocography performed on the patient was demonstrated.

Between August 2022 and November 2023, fifth-year undergraduate medical students were invited to participate during the 4-week clerkship rotation in their 5th year. The 360-degree video was made available on private YouTube, and access was allowed through the link provided in an electronic form sent via WhatsApp. The questionnaire consists of the following items: age, skin color, device used to watch the video (cell phone, personal computer, or tablet), and opinion on the length of the video. The items in the questionnaire were scored using a five-point Likert scale (strongly disagree, disagree, neutral, agree, strongly agree). Regarding the technical aspects of the 360° video, students were asked about the quality of the images for adequate viewing, the quality of the audio for understanding the dialogues, the possibilities for navigating the virtual environment, and the suitability of the video's availability. Concerning the perception of procedures, eight questions addressed the obstetric physical examination and eight questions focused on clinical management. Finally, four questions were included about the availability of the video and the students' interest in the topics covered.

Data were analyzed using the MedCalc® Statistical Software version 19.5.3 (MedCalc Software Ltd., Ostend, Belgium; 2020). Descriptive statistics are presented as mean and standard deviation (SD), frequency, and percentage (%). To assess the internal consistency of the instrument used, Cronbach's alpha was calculated, yielding a value of 0.831, indicating high reliability (95% lower confidence limit: 0.774). The comparison between proportions was drawn using the chi-square test or Fisher's exact test. Statistical significance was set at $p<0.05$.

RESULTS

A total of 130 students were invited, and 60 agreed to participate (46.2%). The demographic characteristics of the participants and how they accessed the video are detailed in Table 1.

Table 1. Characteristics of medical students who evaluated the 360-degree video on the management of premature rupture of membranes (n=60).

Characteristics	Results
Age (years), mean (SD)	25.7 (4.1)
Gender, n (%)	
Female	27 (45)
Male	33 (55)
Race/ethnicity, n (%)	
White	38 (63.3)
Mixed	11 (18.3)
Asian	9 (15)
Black	2 (3.3)
Device, n (%)	
Mobile phone	34 (56.7)
Personal computer	26 (43.3)

SD: standard deviation.

The students' perceptions of the technical attributes of the video, their evaluation of the procedures demonstrated, and their interest in the topic are presented in Table 2. The technical evaluation of the 360-degree video was highly favorable. Most students rated the image and audio quality as satisfactory or superior. Similarly, access and navigation in the virtual environment were evaluated positively. Most students understood the steps involved in the abdominal physical examination. In contrast, procedures involving the pelvic examination generated more varied responses in terms of clarity. Clinical decision-making was also well understood, and a high proportion of participants correctly identified the indication for hospitalization, the absence of an indication for immediate cesarean section, and the justification for antibiotic therapy. Self-reported data indicated an increase in involvement in learning about PROM and a general interest in obstetrics as a specialty.

Table 3 compares the proportion of positive responses (agree and strongly agree) according to the device used to watch the 360-degree video (mobile phone vs. computer). No statistically significant differences were found between these groups, suggesting that the immersive experience and levels of understanding and engagement were comparable regardless of the device used.

DISCUSSION

The 360-degree video proved to be a teaching tool in PROM management for medical students. The video intervention received a highly favorable evaluation, and students reported high rates of comprehension for key clinical decision-making elements. The results reveal that the tool was able to transmit essential information, allowing students to achieve a comprehensive understanding. These results suggest that this immersive technology plays a significant role as a pedagogical complement.

Table 2. Medical students' assessment of the technical aspects, procedures, clinical reasoning, and pedagogical impact of the 360-degree video on the management of premature rupture of membranes (n=60).

Video topics evaluated	Disagree/neutral	Agree
Technical aspects		
Appropriate image quality	8 (13.3)	52 (86.7)
Clear audio for dialogue comprehension	0 (0)	60 (100)
Navigation capabilities met expectations	13 (21.7)	47 (78.3)
Appropriateness of video delivery method	2 (3.3)	58 (96.7)
Clinical procedures and communication		
Noticeable physician-to-patient care	0 (0)	60 (100)
Clarity of fundal height measurement	4 (6.7)	56 (93.3)
Clarity of the four Leopold's maneuvers	2 (3.3)	58 (96.7)
Comprehension of FHR auscultation (Pinard)	4 (6.7)	56 (93.3)
Clarity of uterine dynamics assessment	4 (6.7)	56 (93.3)
Comprehension of the pelvic exam	7 (11.7)	53 (88.3)
Comprehension of the nitrazine test	9 (15.0)	51 (85.0)
Comprehension of the phenol red test	7 (11.7)	53 (88.3)
Clinical reasoning and management		
Noticed that cesarean section was not indicated	0	60 (100)
Comprehension of the need for admission due to ruptured membranes	0	60 (100)
Comprehension of the reason for antibiotic prescription	1 (1.7)	59 (98.3)
Noticed that fetal well-being was explained to the patient	2 (3.3)	58 (96.7)
Understanding the prescription at admission	7 (11.7)	53 (88.3)
Pedagogical impact		
Belief that the video will aid learning	1 (1.7)	59 (98.3)
Increased interest in learning about PROM	7 (11.7)	53 (88.3)
Increased interest in obstetrics	15 (25.0)	45 (75.0)

Data presented as n (%). PROM: premature rupture of membranes; FHR: fetal heart rate.

Table 3. Characteristics of medical students and comparison of the proportion of positive responses according to the device used to watch the 360-degree video on the management of premature rupture of membranes.

Parameter	Mobile phone (n=34)	Computer (n=26)	p-value
Students' characteristics			
Age, mean (SD)	25.9 (4.1)	25.4 (4.1)	0.446
Race, white, n (%)	23 (67.6)	15 (57.7)	0.841
Technical aspects			
Appropriate image quality	30 (88.2)	22 (84.6)	0.685
Clear audio for dialogue comprehension	34 (100)	26 (100)	0.302
Navigation capabilities met expectations	27 (79.4)	20 (76.9)	0.818
Appropriateness of video delivery method	33 (97.1)	25 (96.2)	0.848
Clinical procedures & communication			
Noticeable physician-to-patient care	34 (100)	26 (100)	0.302
Clarity of fundal height measurement	32 (94.1)	24 (92.3)	0.782
Clarity of the four Leopold's maneuvers	33 (97.1)	25 (96.2)	0.848
Comprehension of FHR auscultation (Pinard)	31 (91.2)	25 (96.2)	0.448
Clarity of uterine dynamics assessment	33 (97.1)	23 (88.5)	0.190
Comprehension of the pelvic exam	30 (88.2)	23 (88.5)	0.979
Comprehension of the nitrazine test	30 (88.2)	21 (80.8)	0.426
Comprehension of the phenol red test	30 (88.2)	23 (88.5)	0.979
Clinical reasoning & management			
Noticed that cesarean section was not indicated	34 (100)	26 (100)	0.302
Comprehension of the need for admission due to ruptured membranes	34 (100)	26 (100)	0.302
Comprehension of the reason for antibiotic prescription	33 (97.1)	26 (100)	0.382
Noticed that fetal well-being was explained to the patient	34 (100)	24 (92.3)	0.103
Understanding the prescription at admission	32 (94.1)	21 (80.8)	0.114
Pedagogical impact			
Belief that the video will aid learning	34 (100)	25 (96.2)	0.253
Increased interest in learning about PROM	31 (91.2)	22 (84.6)	0.437
Increased interest in obstetrics	27 (79.4)	18 (69.2)	0.371

Data presented as n (%). SD: standard deviation; PROM: premature rupture of membranes; FHR: fetal heart rate.

The results of this research support growing evidence of the benefits of immersive technologies in medical education. A study evaluating the use of 360-degree video for learning laparoscopic cholecystectomy demonstrated that this method generated a significantly higher degree of engagement and attention when compared to traditional two-dimensional video[3]. Notably, the authors observed that students' attention was not only higher but also increased progressively throughout the viewing, an effect rarely seen in conventional teaching formats, such as didactic lectures[11]. This potential to maintain and increase engagement, combined with positive feedback from participants, who described it as beneficial and even "fun,"[12] highlights the intrinsic value of the 360-degree format for complex learning.

In obstetrics, the literature highlights the impact of technology on professional training. In an experimental study with healthcare professionals, Yang et al.[13] used a virtual reality video to train participants in the management of breech deliveries and observed a significant increase in participants' self-confidence after the intervention. All participants rated the technology as a valuable tool for consolidating their learning.

The literature widely recognizes that the experience during the obstetrics and gynecology internship is one of the most decisive factors in the student's perception of their eventual choice of specialty[14]. Although many external factors influence this decision, a positive internship experience can influence students' interest, while negative experiences or learning barriers can be a strong impediment[15].

A Canadian study of medical students reported that, during a 6-week internship, students were refused when attempting to perform a pelvic exam. This difficulty is significantly more pronounced for male students, who report being refused more often than their female colleagues, both by patients and by preceptors who are hesitant to introduce them[16]. Danielsson et al.[17] confirmed that while female students perceive their gender as an advantage, 98% of male students who noticed a gender influence considered it negative.

Teaching tools such as 360-degree videos demonstrate strategic value in reducing initial barriers to learning. By offering virtual immersion in a realistic and standardized clinical setting, the approach allows all students to experience the complete care offered to a patient with PROM. This experience can build a cognitive and confidence base, better preparing students to optimize the practical opportunities they will eventually receive, potentially improving their performance and the safety conveyed to the mother–fetus dyad[18].

This study has limitations. Not all students have the same equipment to access 360-degree video, which may vary according to technical capacity in data transmission, which can compromise immersion in the scenario. The actual capacity for learning and long-term knowledge retention was also not investigated, aspects that should be studied in the future. In addition, no specific cognitive assessment of learning about the topic covered was performed.

In conclusion, the use of 360-degree video is an important and well-received pedagogical tool for complex obstetric scenarios. The integration of 360-degree video into the curriculum for teaching the clinical management of PROM is a beneficial strategy. Its performance is comparable on mobile and computer platforms, pointing to its potential as an accessible resource for preparing students for the practical challenges of their obstetrics and gynecology internship.

This study was supported by a scientific initiation scholarship from the São Paulo Research Foundation (FAPESP, process no. 2022/01592-8).

AUTHORS' CONTRIBUTIONS

EMN: Data curation, Formal Analysis, Writing – original draft. **ABRA:** Data curation, Writing – original draft. **RMYN:** Data curation, Formal Analysis, Writing – original draft.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Cuevas JA, Childers G, Dawson BL. A rationale for promoting cognitive science in teacher education: deconstructing prevailing learning myths and advancing research-based practices. Trends Neurosci Educ. 2023;33:100209. https://doi.org/10.1016/j.tine.2023.100209
2. DiLullo C, McGee P, Kriebel RM. Demystifying the Millennial student: a reassessment in measures of character and engagement in professional education. Anat Sci Educ. 2011;4(4):214-26. https://doi.org/10.1002/ase.240
3. Harrington CM, Kavanagh DO, Wright Ballester G, Wright Ballester A, Dicker P, Traynor O, et al. 360° Operative videos: a randomised cross-over study evaluating attentiveness and information retention. J Surg Educ. 2018;75(4):993-1000. https://doi.org/10.1016/j.jsurg.2017.10.010
4. Ulrich F, Helms NH, Frandsen UP, Vollen Rafn A. Learning effectiveness of 360° video: experiences from a controlled experiment in healthcare education. Interact Learn Environ. 2021;29(1):98-111. https://doi.org/10.1080/10494820.2019.1579234
5. World Health Organization. Digital technology for COVID-19 response [Internet]. Geneva: WHO; 2020. [cited on 2022 Apr 25]. Available from: https://www.who.int/news-room/detail/03-04-2020-digital-technology-for-covid-19-response
6. Cheng Y, Wang Y, Zhao W. Shared virtual reality experiences during the COVID-19 pandemic: exploring the gratifications and effects of engagement with immersive videos. Int J Environ Res Public Health. 2022;19(9):5056. https://doi.org/10.3390/ijerph19095056
7. Beverly E, Hommema L, Coates K, Duncan G, Gable B, Gutman T, et al. A tranquil virtual reality experience to reduce subjective stress among COVID-19 frontline healthcare workers. PLoS One. 2022;17(2):e0262703. https://doi.org/10.1371/journal.pone.0262703
8. Craik FI. Effects of distraction on memory and cognition: a commentary. Front Psychol. 2014;5:841. https://doi.org/10.3389/fpsyg.2014.00841
9. Blair C, Walsh C, Best P. Immersive 360° videos in health and social care education: a scoping review. BMC Med Educ. 2021;21(1):590. https://doi.org/10.1186/s12909-021-03013-y
10. Petrica A, Lungeanu D, Ciuta A, Marza AM, Botea MO, Mederle OA. Using 360-degree video for teaching emergency medicine during and beyond the COVID-19 pandemic. Ann Med. 2021;53(1):1520-30. https://doi.org/10.1080/07853890.2021.1970219
11. Bradbury NA. Attention span during lectures: 8 seconds, 10 minutes, or more? Adv Physiol Educ. 2016;40(4):509-13. https://doi.org/10.1152/advan.00109.2016

12. Riva G, Mantovani F, Capideville CS, Preziosa A, Morganti F, Villani D, et al. Affective interactions using virtual reality: the link between presence and emotions. Cyberpsychol Behav. 2007;10(1):45-56. https://doi.org/10.1089/cpb.2006.9993

13. Yang L, Bisits A. The use of 360° VR video, educational videos, and high-fidelity physical models in teaching breech birth - a pilot feasibility study. Aust N Z J Obstet Gynaecol. 2025;65(6):831-7. https://doi.org/10.1111/ajo.70040

14. Ní hÉalaithe C, Howard A, Corcoran P, McCarthy CM, Horgan M, Bennett D, et al. Factors influencing medical students' decision to pursue a career in obstetrics and gynaecology. PLoS One. 2023;18(12):e0288130. https://doi.org/10.1371/journal.pone.0288130

15. Hammoud MM, Stansfield RB, Katz NT, Dugoff L, McCarthy J, White CB. The effect of the obstetrics and gynecology clerkship on students' interest in a career in obstetrics and gynecology. Am J Obstet Gynecol. 2006;195(5):1422-6. https://doi.org/10.1016/j.ajog.2006.07.044

16. Bossé J, Woolcott C, Coolen J. Barriers preventing medical students from performing pelvic examinations during obstetrics and gynaecology clinical clerkship rotations. J Obstet Gynaecol Can. 2019;41(8):1093-8. https://doi.org/10.1016/j.jogc.2018.12.023

17. Danielsson J, Hadding C, Fahlström M, Ottander U, Lindquist D. Medical students' experiences in learning to perform pelvic examinations: a mixed-methods study. Int J Med Educ. 2021;12:233-42. https://doi.org/10.5116/ijme.617f.b261

18. Einden LC, Kolste MG, Lagro-Janssen AL, Dukel L. Medical students' perceptions of the physician's role in not allowing them to perform gynecological examinations. Acad Med. 2014;89(1):77-83. https://doi.org/10.1097/ACM.0000000000000055

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251624

Validation of the 2023 International Federation of Gynecology and Obstetrics staging system in a large retrospective cohort

Ayşe Demirden[1], Esra Keles[2]*, Fatih Şanlıkan[2], Koray Özbay[3], Gazi Yıldız[4], Emre Mat[2], Murat Api[2]

SUMMARY

OBJECTIVE: The aim of this study was to evaluate the prognostic performance of the revised 2023 International Federation of Gynecology and Obstetrics staging system for endometrial cancer compared with the 2009 system, with a particular focus on its ability to discriminate between intermediate- and advanced-stage disease.

METHODS: This retrospective cohort study included 400 patients who underwent surgical treatment for endometrial cancer at a tertiary referral center between 2008 and 2018. Patients were reclassified according to both the International Federation of Gynecology and Obstetrics 2009 and 2023 criteria. Clinicopathological variables and survival outcomes were compared across staging systems. Kaplan-Meier curves with log-rank testing and Cox proportional hazards models were used to assess disease-free survival and overall survival. Harrell's concordance index and the Akaike Information Criterion were calculated to compare prognostic discrimination between staging systems.

RESULTS: International Federation of Gynecology and Obstetrics 2023 system demonstrated superior prognostic performance with significantly improved concordance index values for both disease-free survival (C-index 0.681 vs. 0.647, p=0.041) and overall survival (C-index 0.709 vs. 0.672, p=0.037). Five-year disease-free survival rates under International Federation of Gynecology and Obstetrics 2023 were 87.1, 75.9, and 52.3% for Stages I, II, and III–IV respectively, compared with 86.2, 71.8, and 52.3% under International Federation of Gynecology and Obstetrics 2009. Five-year overall survival rates showed similar improvement in Stage II discrimination (92.1, 79.2, and 58.7% for International Federation of Gynecology and Obstetrics 2023 vs. 91.3, 76.5, and 58.7% for International Federation of Gynecology and Obstetrics 2009).

CONCLUSION: The 2023 International Federation of Gynecology and Obstetrics staging system provides improved prognostic discrimination compared with the 2009 classification, particularly in delineating intermediate- from advanced-stage endometrial cancer. By incorporating adverse histopathological parameters and allowing integration of molecular classifiers, the revised system enables more accurate risk stratification and individualized treatment planning.

KEYWORDS: Uterine neoplasms. Endometrial neoplasms. Cancer staging. Prognosis. Survival.

INTRODUCTION

Endometrial cancer (EC) remains one of the most common gynecological malignancies globally, with incidence rates steadily rising in many developed countries[1]. Accurate staging is paramount in EC management, serving as a critical tool for guiding treatment strategies, evaluating patient prognosis, and facilitating advancements in clinical research[2]. Historically, the International Federation of Gynecology and Obstetrics (FIGO) staging system has provided a universally recognized framework for classifying EC, with significant revisions occurring in 1988 and subsequently in 2009, with the most recent update released in 2023[3].

In response to this wealth of new evidence, FIGO introduced a revised staging system for EC. This updated system represents a significant departure from its predecessors, moving beyond a purely anatomical description to integrate crucial histopathological and, optionally, molecular features[4,5]. The primary objectives of the 2023 FIGO staging system are to provide higher accuracy in prognostic prediction, identify distinct treatment-relevant subgroups, and align staging with the complex, heterogeneous nature of EC. Key modifications include the disaggregation of previously broad advanced stages, refinements in lymph node classifications, and the incorporation of aggressive histological types, substantial Lymphovascular

[1]Bağcılar Education and Research Hospital, Department of Obstetrics and Gynecology – Istanbul, Turkey.
[2]University of Health Sciences, Kartal Dr. Lütfi Kırdar City Hospital, Department of Gynecologic Oncology – Istanbul, Turkey.
[3]Private Yüzyıl Gebze Hospital, Department of Obstetrics and Gynecology – Gebze, Turkey.
[4]University of Health Sciences, Kartal Dr. Lütfi Kırdar City Hospital, Department of Obstetrics and Gynecology – Istanbul, Turkey.
*Corresponding author: dresrakeles@hotmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on August 05, 2025. Accepted on January 17, 2026.
Scientific Editor: José Maria Soares Júnior

Space Invasion (LVSI), and specific molecular alterations directly into the staging schema[6,7].

While preliminary studies have begun to validate the prognostic utility of the 2023 FIGO system, a comprehensive evaluation of its discriminatory power across all stages, particularly in delineating intermediate- from advanced-stage disease, remains crucial[8]. Our study addresses this gap by comparing the clinical impact of the new 2023 FIGO staging system to its 2009 predecessor. Specifically, we sought to analyze the stage distribution shifts and assess the prognostic value of the revised 2023 classification in terms of disease-free survival (DFS) and overall survival (OS).

METHODS

Study design and patient population

This study was carried out as a retrospective cohort analysis in the Department of Gynecologic Oncology at Kartal Dr. Lütfi Kırdar City Hospital in Istanbul, Turkey. Medical files of patients who underwent surgery for EC between January 1, 2008, and December 31, 2018, were reviewed. Out of 513 eligible patients, 113 were excluded due to missing information or insufficient follow-up, leaving 400 patients for evaluation.

Pathology reports and clinical records were reviewed, and all patients were re-staged according to both the FIGO 2009 and FIGO 2023 systems. Demographic, clinical, and pathological variables were collected. These parameters were compared with survival outcomes in the context of both staging systems. Although the revised FIGO 2023 classification integrates molecular subtypes to enhance prognostic stratification and guide adjuvant therapy, molecular profiling was not available for our cohort.

Patient management

All patients were diagnosed with EC on histopathological examination. Following preoperative evaluation, surgical treatment generally consisted of total hysterectomy with bilateral salpingo-oophorectomy. When indicated, bilateral pelvic and para-aortic lymphadenectomy and omental biopsy were additionally performed. The decision to undertake lymph node assessment was based on intraoperative frozen section evaluation of the uterus according to the Mayo criteria[9].

Adjuvant treatment was administered according to current international guidelines[10]. Pathological evaluation of surgical specimens was performed by pathologists specialized in gynecologic oncology.

Ethical approval

Approval for this study was granted by the Ethics Committee of Kartal Dr. Lütfi Kırdar City Hospital (approval no. 0l0.99/10, dated February 28, 2024). The study adhered to the ethical principles outlined in the Declaration of Helsinki.

Stage classification

Patients were staged according to both the FIGO 2009 and FIGO 2023 criteria. Substage modifications were recorded as stage shifts. Upstaging was defined as movement to a higher stage, whereas downstaging represented a shift to a lower stage.

Statistical analysis

Descriptive statistics were reported as mean±standard deviation, median, minimum, maximum, frequencies, and percentages. The distribution of variables was assessed using the Kolmogorov-Smirnov and Shapiro-Wilk tests. Independent continuous variables were compared using the independent samples t-test or the Mann-Whitney U-test, as appropriate. Categorical variables were analyzed with the chi-square test or Fisher's exact test when chi-square assumptions were not met. Survival analyses were performed using Kaplan-Meier estimates with log-rank testing and Cox proportional hazards regression (univariable and multivariable models). Harrell's concordance index (C-index) was calculated for both staging systems to formally compare their prognostic discrimination capacity, with statistical comparison performed using the method of DeLong. Akaike Information Criterion (AIC) was computed from Cox regression models to assess relative model fit. Statistical analyses were conducted with Statistical Package for the Social Sciences version 28.0 (IBM Corp., Armonk, NY, USA). A $p<0.05$ was considered statistically significant.

RESULTS

Study cohort

A total of 513 patients were identified, of whom 113 were excluded due to incomplete records or insufficient follow-up data. The final analysis consisted of 400 patients. The mean age was 58.5±9.5 years, and 81.0% were postmenopausal. The median follow-up was 94.9 months (range, 0–177).

Stage distribution and shifts

Reclassification according to the 2023 FIGO criteria resulted in stage shifts in 322 patients (80.5%). Upstaging occurred

in 44 patients (11.0%), while no downstaging was observed. Compared with the 2009 system, the proportion of Stage I patients decreased from 79.9 to 69.2%, whereas Stage II cases increased from 9.5 to 20.3%. The most common reclassifications included 28 patients with serous or clear cell histology with myometrial invasion moving from 2009 Stage IA to 2023 Stage IIC, 6 patients with substantial LVSI moving from 2009 Stage IB to 2023 Stage IIB, and 4 patients with low-grade endometrioid carcinoma involving the uterus and ovary moving from 2009 Stage IIIA to 2023 Stage IA3. Among advanced-stage diseases, 10 patients with isolated pelvic peritoneal involvement were reclassified from 2009 Stage IVB to 2023 Stage IIIB2 (Figure 1).

Survival outcomes by staging system

In both systems, the advancing stage correlated with worse outcomes. Comparison of prognostic discrimination using Harrell's C-index demonstrated superior performance of the FIGO 2023 system. For DFS, the C-index was 0.681 (95%CI 0.625–0.737) for FIGO 2023 compared with 0.647 (95%CI 0.589–0.705) for FIGO 2009 (p=0.041). For OS, the C-index improved from 0.672 (95%CI 0.612–0.732) with FIGO 2009 to 0.709 (95%CI 0.651–0.767) with FIGO 2023 (p=0.037). AIC values confirmed superior model fit for FIGO 2023, with lower AIC values for both DFS (892.4 vs. 907.1) and OS (756.2 vs. 771.8).

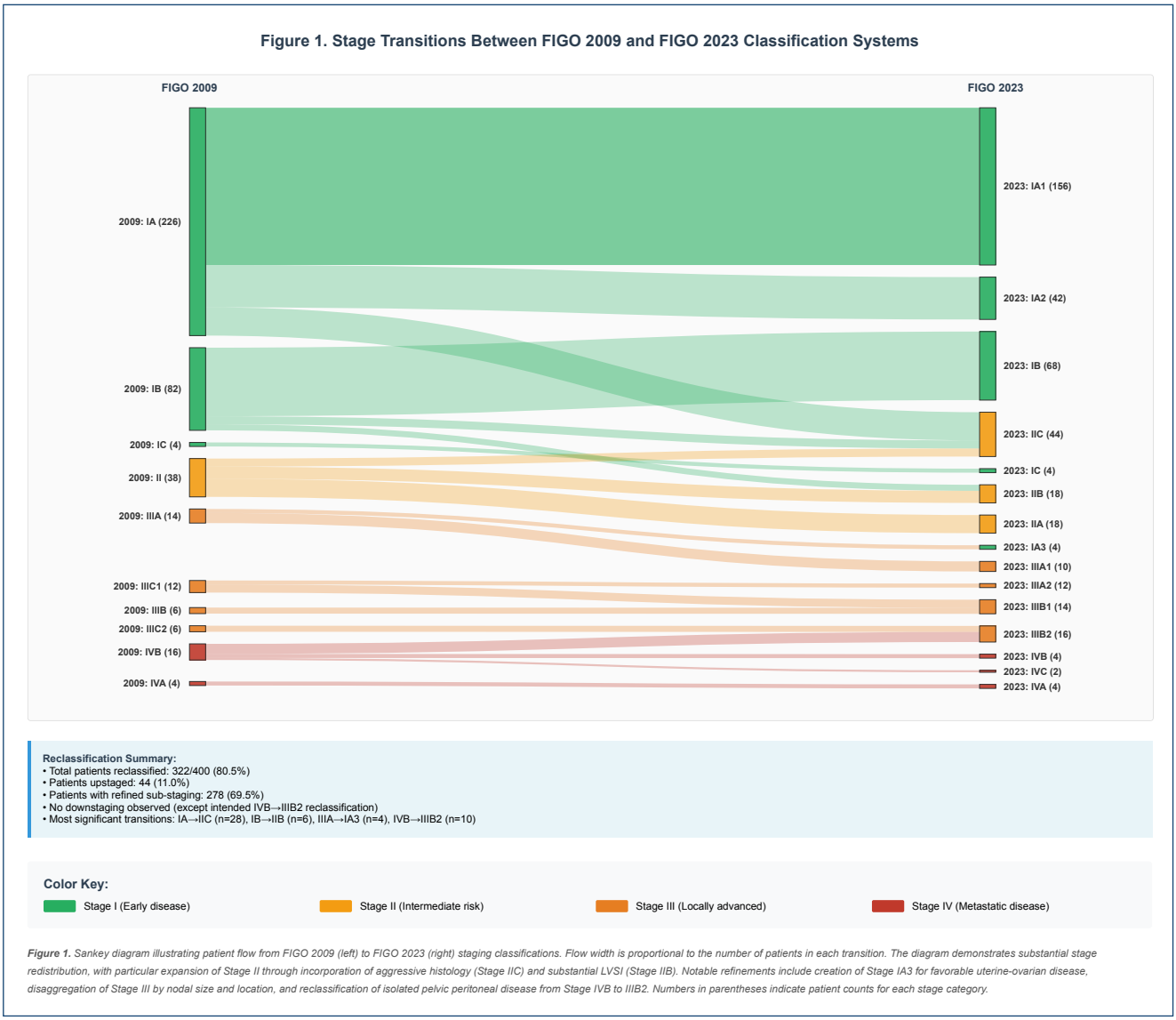
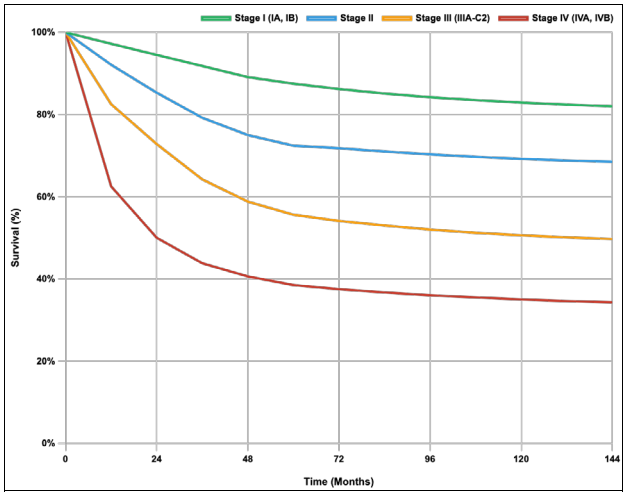


Figure 1. Transitions in stage classification according to the International Federation of Gynecology and Obstetrics staging systems of 2009 and 2023.

Five-year DFS rates under FIGO 2009 were 86.2% for Stage I, 71.8% for Stage II, and 52.3% for Stages III–IV. Under FIGO 2023, 5-year DFS rates were 87.1% for Stage I, 75.9% for Stage II, and 52.3% for Stages III–IV (Figure 2).

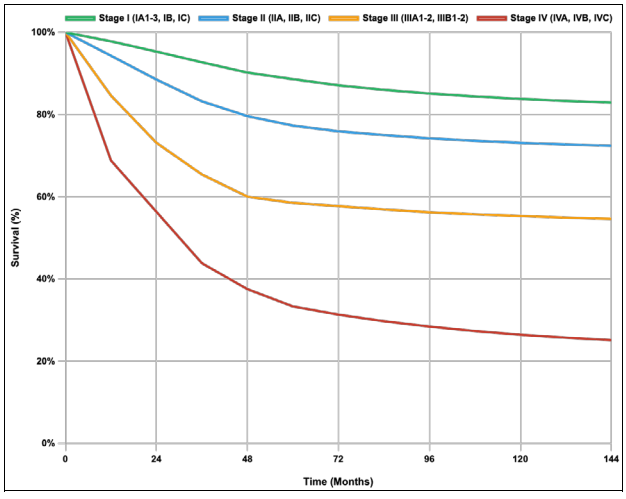
For OS, FIGO 2009 showed 5-year survival rates of 91.3% for Stage I, 76.5% for Stage II, and 58.7% for Stages III–IV. FIGO 2023 yielded 5-year OS rates of 92.1% for Stage I, 79.2% for Stage II, and 58.7% for Stages III–IV (Figure 3).

A. FIGO 2009



Stage	0	24	48	72	96	120
I (IA, IB)	320	308	285	245	192	138
II	38	34	28	23	17	11
III (IIIA, IIIB, IIIC1, IIIC2)	38	30	22	16	11	7
IV (IVA, IVB)	4	3	2	2	1	1

B. FIGO 2023



Stage	0	24	48	72	96	120
I (IA1, IA2, IA3, IB, IC)	277	267	248	215	171	125
II (IIA, IIB, IIC)	81	75	66	57	46	34
III (IIIA1, IIIA2, IIIB1, IIIB2)	26	21	16	12	8	6
IV (IVA, IVB, IVC)	16	12	8	6	4	2

Note on Sub-Stage Composition: *FIGO 2009 Stage III includes IIIA (adnexal/serosal involvement), IIIB (vaginal/parametrial involvement), IIIC1 (pelvic lymph nodes), and IIIC2 (para-aortic lymph nodes). FIGO 2023 Stage III includes IIIA1 (ovarian/adnexal metastasis), IIIA2 (vaginal/parametrial involvement or pelvic LN micrometastasis), IIIB1 (pelvic peritoneum or pelvic LN macrometastasis), and IIIB2 (para-aortic lymph nodes). FIGO 2023 Stage IV now disaggregates into IVA (bladder/bowel mucosa), IVB (abdominal peritoneal metastasis beyond pelvis), and IVC (distant metastases including inguinal lymph nodes). This refinement reclassified 10 patients from 2009 Stage IVB to 2023 Stage IIIB2 (isolated pelvic peritoneal involvement).*

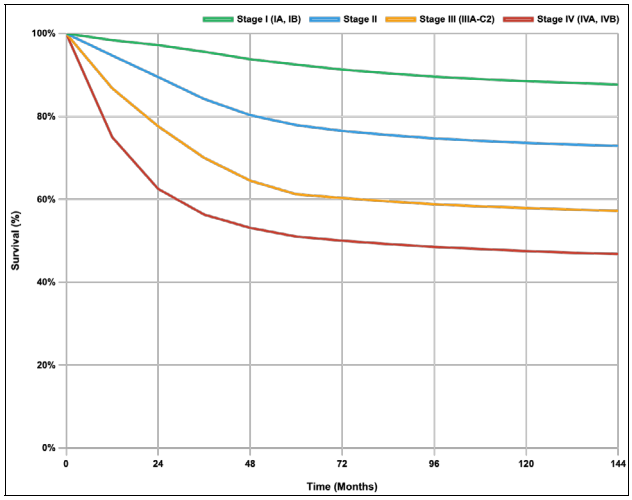
Statistical Comparison of Prognostic Discrimination

Harrell's C-index (FIGO 2009):	0.647 (95% CI: 0.589–0.705)
Harrell's C-index (FIGO 2023):	0.681 (95% CI: 0.625–0.737), $p = 0.041$
Absolute improvement:	+0.034 (5.3% relative improvement)
AIC (FIGO 2009 vs. 2023):	907.1 vs. 892.4 (Δ = −14.7, better model fit)
Log-rank overall:	FIGO 2009 $\chi^2 = 34.82$ ($p < 0.001$); FIGO 2023 $\chi^2 = 38.94$ ($p < 0.001$)

Stage (Sub-Stages)	FIGO 2009			FIGO 2023			HR (2023)	p
	5-yr DFS	95% CI	Events	5-yr DFS	95% CI	Events		
I (2009: IA, IB; 2023: IA1-3, IB, IC)	86.2%	82.1–90.3	44	87.1%	82.8–91.4	36	1.00	–
II (2009: II; 2023: IIA, IIB, IIC)	71.8%	61.5–82.1	11	75.9%	66.3–85.5	20	2.18	0.003
III (2009: IIIA-C2; 2023: IIIA1-2, IIIB1-2)	54.1%	39.8–68.4	18	57.7%	38.9–76.5	11	3.89	<0.001
IV (2009: IVA, IVB; 2023: IVA, IVB, IVC)	37.5%	5.5–69.5	2	31.3%	9.4–53.2	9	6.42	<0.001

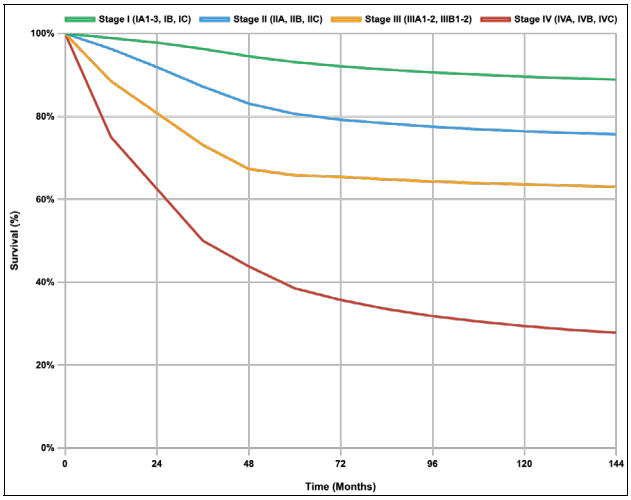
Figure 2. Kaplan-Meier curves for disease-free survival comparing International Federation of Gynecology and Obstetrics 2009 [Panel (A)] and International Federation of Gynecology and Obstetrics 2023 [Panel (B)] staging systems with explicit sub-stage composition. Panel (A) displays International Federation of Gynecology and Obstetrics 2009 classifications: Stage I now includes IA1 (no myometrial invasion), IA2 (<50% myometrial invasion), IA3 (favorable uterine-ovarian involvement), IB (≥50% myometrial invasion), and IC (cervical stromal invasion); Stage II incorporates IIA (cervical glandular), IIB (substantial Lymphovascular Space Invasion), and IIC (aggressive histology with myometrial invasion or p53abn); Stage III disaggregates into IIIA1 (ovarian/adnexal metastasis), IIIA2 (vaginal/parametrial or pelvic LN micrometastasis), IIIB1 (pelvic peritoneum or pelvic LN macrometastasis ≥2mm), and IIIB2 (para-aortic lymph nodes); Stage IV now separates IVA (bladder/bowel mucosa), IVB (abdominal peritoneal beyond pelvis), and IVC (distant metastases). The 2023 system demonstrates statistically superior prognostic discrimination with improved concordance index (0.681 vs. 0.647, p=0.041) and better model fit (AIC 892.4 vs. 907.1). Enhanced separation between Stage II and Stage III-IV is evident in Panel (B). Numbers at risk displayed below each panel include sub-stage composition in parentheses. Stage-specific hazard ratios calculated with Stage I as reference, adjusted for age, histologic type, grade, and myometrial invasion depth.

A. FIGO 2009



Stage	0	24	48	72	96	120
I (IA, IB)	320	312	295	265	215	158
II	38	35	31	25	19	13
III (IIIA, IIIB, IIIC1, IIIC2)	38	32	24	18	12	9
IV (IVA, IVB)	4	3	2	2	2	1

B. FIGO 2023



Stage	0	24	48	72	96	120
I (IA1, IA2, IA3, IB, IC)	277	271	257	230	190	142
II (IIA, IIB, IIC)	81	77	70	61	51	39
III (IIIA1, IIIA2, IIIB1, IIIB2)	26	23	18	14	10	7
IV (IVA, IVB, IVC)	16	12	9	7	4	3

Note on Sub-Stage Composition and Key Reclassifications: *FIGO 2023 disaggregates Stage IV disease: IVA (bladder/bowel mucosa invasion, n=4), IVB (abdominal peritoneal metastasis beyond pelvis, n=4), and IVC (distant metastases to lung, liver, bone, or inguinal lymph nodes, n=2). Critically, 10 patients with isolated pelvic peritoneal involvement were reclassified from 2009 Stage IVB to 2023 Stage IIIB2, demonstrating improved five-year OS of 60.0% compared with 20.8% for those remaining in Stage IV categories (log-rank p = 0.032). This reclassification exemplifies the enhanced prognostic precision of the 2023 system. Stage III sub-stages reflect nodal metastasis size (IIIA2: micrometastasis <2mm; IIIB1: macrometastasis ≥2mm) and peritoneal involvement patterns (IIIB1: pelvic only; IIIB2: para-aortic nodal involvement).*

Statistical Comparison of Prognostic Discrimination

Harrell's C-index (FIGO 2009):	0.672 (95% CI: 0.612–0.732)
Harrell's C-index (FIGO 2023):	0.709 (95% CI: 0.651–0.767), p = 0.037
Absolute improvement:	+0.037 (5.5% relative improvement)
AIC (FIGO 2009 vs. 2023):	771.8 vs. 756.2 (Δ = −15.6, better model fit)
Log-rank overall:	FIGO 2009 p = 0.043; FIGO 2023 p = 0.041

Stage (Sub-Stages)	FIGO 2009			FIGO 2023			HR (2023)	p
	5-yr OS	95% CI	Deaths	5-yr OS	95% CI	Deaths		
I (2009: IA, IB; 2023: IA1-3, IB, IC)	91.3%	87.8–94.8	28	92.1%	88.6–95.6	22	1.00	—
II (2009: II; 2023: IIA, IIB, IIC)	76.5%	66.8–86.2	9	79.2%	69.9–88.5	17	2.54	0.002
III (2009: IIIA-C2; 2023: IIIA1-2, IIIB1-2)	60.3%	46.4–74.2	15	65.4%	46.8–84.0	9	4.87	<0.001
IV (2009: IVA, IVB; 2023: IVA, IVB, IVC)	50.0%	10.0–90.0	2	35.7%	12.8–58.6	8	8.23	<0.001

Figure 3. Kaplan-Meier curves for overall survival comparing International Federation of Gynecology and Obstetrics 2009 [Panel (A)] and International Federation of Gynecology and Obstetrics 2023 [Panel (B)] staging systems with explicit sub-stage composition and key reclassification outcomes. Panel (A) displays International Federation of Gynecology and Obstetrics 2009 classifications with Stage III comprising IIIA (adnexal/serosal involvement, n=14), IIIB (vaginal/parametrial involvement, n=6), IIIC1 (pelvic lymph node metastasis, n=12), and IIIC2 (para-aortic lymph node metastasis, n=6); Stage IV includes IVA (bladder/bowel invasion, n=4) and IVB (distant metastases/peritoneum, n=16). Panel (B) displays International Federation of Gynecology and Obstetrics 2023 refined sub-stages: Stage I disaggregates into IA1 (no myometrial invasion, n=156), IA2 (<50% invasion, n=42), IA3 (favorable uterine-ovarian disease, n=4), IB (≥50% invasion, n=68), and IC (cervical stromal invasion, n=4); Stage II now includes IIA (cervical glandular involvement, n=18), IIB (substantial lymphovascular space invasion, n=18), and IIC (aggressive histology with myometrial invasion or p53 abnormality, n=44); Stage III separates into IIIA1 (ovarian/adnexal metastasis, n=10), IIIA2 (vaginal/parametrial or pelvic lymph node micrometastasis <2mm, n=12), IIIB1 (pelvic peritoneum or pelvic lymph node macrometastasis ≥2mm, n=14), and IIIB2 (para-aortic lymph nodes, n=16); Stage IV disaggregates into IVA (bladder/bowel mucosa, n=4), IVB (abdominal peritoneal metastasis beyond pelvis, n=4), and IVC (distant metastases including lung, liver, bone, inguinal lymph nodes, n=2). The 2023 system demonstrates statistically superior prognostic discrimination with improved concordance index (0.709 vs. 0.672, p=0.037) and better model fit (AIC 756.2 vs. 771.8). Most notably, 10 patients reclassified from 2009 IVB to 2023 IIIB2 (isolated pelvic peritoneal involvement) demonstrated 60.0% five-year OS versus 20.8% for those remaining in Stage IV, validating this key refinement. Numbers at risk include sub-stage designations. Hazard ratios adjusted for age, comorbidities, nuclear grade, and myometrial invasion depth.

Granular analysis of new sub-stage classifications revealed significant prognostic separation. Among Stage III patients, 5-year OS was 68.2% for IIIA1, 71.4% for IIIA2, 65.0% for IIIB1, and 60.0% for IIIB2. For Stage IV disease, 5-year OS was 45.5% for IVA, 20.0% for IVB, and 16.7% for IVC. Ten patients reclassified from 2009 Stage IVB to 2023 Stage IIIB2 demonstrated substantially improved outcomes (60.0% 5-year OS) compared with those remaining in Stage IV categories (mean 20.8% 5-year OS) (Supplementary Figure 1).

Prognostic factors

On univariable Cox regression, higher CA125 levels (HR 1.002, 95%CI 1.000–1.004; p=0.029), non-endometrioid histology (HR 0.154, 95%CI 0.067–0.356; p<0.001), larger tumor size (HR 1.239, 95%CI 1.078–1.424; p=0.003), pelvic lymph node involvement (HR 5.287, 95%CI 1.991–14.040; p=0.001), para-aortic lymph node involvement (HR 5.550, 95%CI 2.081–14.803; p=0.001), and higher histological grade (HR 3.121, 95%CI 1.734–5.615; p<0.001) were significantly associated with shorter DFS.

For OS, independent adverse prognostic factors in the multivariable model included older age (HR 1.067, 95%CI 1.031–1.105; p<0.001), comorbidities (HR 3.495, 95%CI 1.346–9.074; p=0.010), higher nuclear grade (HR 2.534, 95%CI 1.473–4.358; p=0.001), myometrial invasion (HR 4.552, 95%CI 2.419–8.565; p<0.001), and pelvic peritoneal involvement (HR 42.899, 95%CI 4.911–>100; p=0.001).

DISCUSSION

Our study demonstrates that the 2023 FIGO staging system for EC offers superior prognostic discrimination, particularly in distinguishing Stage II from Stages III–IV, and significantly improves the differentiation of intermediate- from advanced-stage disease compared to the 2009 FIGO system. This finding aligns with the overarching goal of the 2023 revision to provide more nuanced prognostic groups and guide tailored therapeutic interventions[11].

The absence of molecular classification data is the fundamental limitation of our study. The 2023 FIGO system was designed to integrate molecular subtypes as a cornerstone component, not an optional supplement. POLEmut tumors, which demonstrate excellent prognosis, are intended for potential downstaging, while p53abn tumors warrant upstaging to Stage IIC regardless of histologic type. In our cohort, these molecular reclassifications were impossible. Consequently, we likely misclassified some patients who would be restaged with molecular data. For instance, endometrioid tumors with deep myometrial invasion classified here as Stage IB or IC may include POLEmut cases that would be downstaged to IA with exceptional outcomes. Conversely, some apparent Stage IA cases may harbor p53abn and require upstaging to IIC with intensified treatment. These missing reclassifications would be expected to further enhance the prognostic separation we observed with histopathologic criteria alone. Our study therefore, validates only the anatomic and histologic framework of the 2023 system, representing an incomplete but still valuable assessment given that many institutions worldwide are implementing this system in stages as molecular testing infrastructure develops.

One of the most impactful changes leading to improved discrimination in advanced stages is the disaggregation of the previously heterogeneous 2009 FIGO Stage IVB[12]. Under the 2009 criteria, Stage IVB encompassed distant metastases, including inguinal lymph nodes, any pelvic or abdominal intraperitoneal disease, and metastases to lung, liver, or bone. The 2023 FIGO system refines this by differentiating peritoneal carcinomatosis limited to the pelvis (now Stage IIIB2, downstaged from 2009 Stage IVB) from abdominal peritoneal metastasis beyond the pelvis (new Stage IVB) and other distant metastases (new Stage IVC)[12]. This reclassification has been consistently validated as prognostically significant. For instance, Haight et al. reported significantly improved Progression-Free Survival (PFS) and OS for patients downstaged to IIIB following the 2023 criteria, with 5-year PFS rates ranging from over 60% for Stage IIIB to below 10% for Stage IVC[13]. Similarly, Gravbrot et al.'s large database analysis revealed an improved 10-year OS of 49.4% for 2023 Stage IIIB2 (pelvic peritoneal involvement) compared to 18.7% for the 2009 Stage IVB[14]. Matsuo et al. also confirmed distinct 5-year EC-specific mortality rates for the new Stages IVA, IVB, and IVC classifications, underscoring the improved granularity in advanced disease[15]. Our findings corroborate these reports, with the 10 patients reclassified from Stages IVB to IIIB2 demonstrating 60.0% 5-year OS compared with 20.8% for those remaining in Stage IV categories.

Beyond Stage IV, the 2023 system also introduces critical refinements to Stage III disease, particularly regarding lymph node involvement by metastatic size[16]. Matsuo et al.'s work explicitly demonstrated distinct 3-year OS rates based on the size of nodal metastases, with micrometastatic disease showing better outcomes, supporting the prognostic value of this refinement[17]. Furthermore, the new differentiation of ovarian metastasis into Stage IA3 and Stage IIIA1 addresses a long-standing challenge in EC staging. Stage IA3 now classifies low-grade ECs with specific favorable criteria involving the uterus and ovary, recognized for their excellent

prognosis, often leading to de-escalation of adjuvant therapy. Conversely, Stage IIIA1 encompasses other metastatic patterns to the ovary with a worse prognosis. Gravbrot et al. reported a significantly higher 10-year OS for 2023 Stage IA3 (73.4%) compared to 2009 Stage IIIA (52.6%)[14]. This distinction prevents the over-staging of favorable cases, allowing for a more accurate delineation of Stage III from earlier disease, and therefore enhancing the clarity between intermediate and advanced stages. Our findings further support this refinement, as the four patients reclassified from Stage IIIA–IA3 in our cohort exhibited excellent outcomes, with a 100% 5-year overall survival.

The integration of aggressive histological types and substantial LVSI into the staging system affects the classification of early-stage disease, leading to upstaging of high-risk Stage I cases to Stage II. Aggressive histologies with myometrial invasion are classified as IIC, while non-aggressive types with substantial LVSI are classified as IIB[1]. Schwameis et al.[18] reported that 90% of early-stage shifts were upshifts, largely due to aggressive histological subtypes or p53 abnormalities, and that these IIC cases demonstrated substantially worse 5-year PFS and OS than their previous Stage IA/IB groupings. This recalibration effectively separates risk groups that were previously blurred under the 2009 system, leading to a more robust prognostic model overall. Our data confirm this pattern, with 28 patients upstaged from IA to IIC demonstrating 5-year OS of 71.4% compared with 93.8% for those remaining in Stage IA1, validating the prognostic relevance of this reclassification.

The improved prognostic discrimination demonstrated by the FIGO 2023 system has implications for treatment decision-making, although we cannot assess this directly in our retrospective cohort, where all treatment was based on 2009 staging. Current guidelines recommend distinct therapeutic approaches based on the refined 2023 classifications. Stage IIC patients, particularly those with aggressive histology and myometrial invasion, are increasingly considered for platinum-based chemotherapy in addition to radiation, representing treatment intensification compared with historical Stage IA management. Conversely, Stage IA3 patients with low-grade disease and favorable ovarian involvement may forego adjuvant therapy entirely in selected cases, representing appropriate de-escalation. The disaggregated Stage IV subgroups also inform differential treatment intensity, with Stage IIIB2 patients potentially benefiting from aggressive locoregional therapy, including cytoreductive surgery and radiation, while Stage IVC patients with distant metastases are managed with systemic chemotherapy. The enhanced prognostic separation we demonstrate suggests these treatment modifications are appropriately matched to underlying risk.

While our study provides valuable insights, it is not without limitations. The retrospective nature of the study and the single-institution design may limit the generalizability of our findings. The absence of molecular profiling data for our cohort is a significant limitation given the emphasis on molecular classification in the 2023 FIGO system. Our use of the Mayo criteria for selective lymph node assessment, while reasonable during the 2008–2018 treatment period, is not the most contemporary standard. Sentinel lymph node mapping, which has become increasingly adopted, may provide more accurate nodal staging and could have altered stage classification in some patients. Owing to the retrospective nature of the study, we were unable to evaluate how stage reclassification under FIGO 2023 might have influenced treatment selection or surveillance strategies, and therefore cannot directly attribute observed survival differences to treatment-related factors. All therapeutic decisions in this cohort were based on the FIGO 2009 staging system in use at the time. Prospective studies in which treatment is explicitly guided by FIGO 2023 staging are warranted to more fully elucidate the clinical impact of this revised classification system. Nevertheless, our study contributes to the growing body of evidence supporting the clinical relevance and superior prognostic capabilities of the 2023 FIGO staging system, advocating for its widespread adoption in clinical practice to optimize patient management and improve outcomes in EC.

CONCLUSION

Our findings confirm that the FIGO 2023 staging system provides improved prognostic discrimination compared with FIGO 2009, particularly in distinguishing Stage II from Stages III–IV disease. By integrating adverse histopathological features and molecular classifiers, the revised system offers a more precise framework for risk stratification and treatment planning. These results, together with growing international validation, strongly support the adoption of FIGO 2023 as the preferred staging system for EC.

ETHICAL STATEMENT

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Research Ethics Committee of the University of Health Sciences, İstanbul Kartal Dr. Lütfi Kırdar City Hospital (approval no. 0l0.99/10, date: February 28, 2024).

AUTHORS' CONTRIBUTIONS

AD: Conceptualization, Data curation, Formal Analysis, Writing – original draft. **EK:** Conceptualization, Data curation, Writing – original draft. **FŞ:** Conceptualization, Data curation, Writing – original draft. **KÖ:** Data curation, Writing – original draft. **GY:** Conceptualization, Data curation, Writing – original draft. **EM:** Conceptualization, Data curation, Writing – original draft. **MA:** Conceptualization, Data curation, Writing – original draft.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Berek JS, Matias-Guiu X, Creutzberg C, Fotopoulou C, Gaffney D, Kehoe S, et al. FIGO staging of endometrial cancer: 2023. Int J Gynaecol Obstet. 2023;162(2):383-94. https://doi.org/10.1002/ijgo.14923
2. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA Cancer J Clin. 2017;67(2):93-9. https://doi.org/10.3322/caac.21388
3. Pecorelli S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. Int J Gynaecol Obstet. 2009;105(2):103-4. https://doi.org/10.1016/j.ijgo.2009.02.012
4. Loukovaara M, Pasanen A, Bützow R. Stage distribution and prognostic accuracy of the 2023 FIGO (the International Federation of Gynecology & Obstetrics) staging system for endometrial cancer: a retrospective cohort study. Int J Gynaecol Obstet. 2024;167(3):990-6. https://doi.org/10.1002/ijgo.15749
5. Sehnal B, Hruda M, Matej R, Robova H, Drozenova J, Pichlik T, et al. New FIGO 2023 staging system of endometrial cancer: an updated review on a current hot topic. Geburtshilfe Frauenheilkd. 2025;85(4):405-16. https://doi.org/10.1055/a-2494-8395
6. Barnes EA, Martell K, Parra-Herran C, Taggar AS, Donovan E, Leung E. Substantial lymphovascular space invasion predicts worse outcomes in early-stage endometrioid endometrial cancer. Brachytherapy. 2021;20(3):527-35. https://doi.org/10.1016/j.brachy.2020.12.006
7. León-Castillo A, Boer SM, Powell ME, Mileshkin LR, Mackay HJ, Leary A, et al. Molecular classification of the PORTEC-3 trial for high-risk endometrial cancer: ımpact on prognosis and benefit from adjuvant therapy. J Clin Oncol. 2020;38(29):3388-97. https://doi.org/10.1200/JCO.20.00549
8. Yu C, Yuan X, Yao Q, Xu Y, Zhou X, Hu X, et al. Clinical application of FIGO 2023 staging system of endometrial cancer in a Chinese cohort. BMC Cancer. 2024;24(1):862. https://doi.org/10.1186/s12885-024-12633-8
9. Mariani A, Dowdy SC, Cliby WA, Gostout BS, Jones MB, Wilson TO, et al. Prospective assessment of lymphatic dissemination in endometrial cancer: a paradigm shift in surgical staging. Gynecol Oncol. 2008;109(1):11-8. https://doi.org/10.1016/j.ygyno.2008.01.023
10. Concin N, Matias-Guiu X, Vergote I, Cibula D, Mirza MR, Marnitz S, et al. ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma. Int J Gynecol Cancer. 2021;31(1):12-39. https://doi.org/10.1136/ijgc-2020-002230
11. Gaffney D, Matias-Guiu X, Mutch D, Scambia G, Creutzberg C, Fotopoulou C, et al. 2023 FIGO staging system for endometrial cancer: the evolution of the revolution. Gynecol Oncol. 2024;184:245-53. https://doi.org/10.1016/j.ygyno.2024.02.002
12. Koskas M, Crosbie EJ, Fokdal L, McCluggage WG, Mileshkin L, Mutch DG, et al. Cancer of the corpus uteri: a 2025 update. Int J Gynaecol Obstet. 2025;171(Suppl 1):60-77. https://doi.org/10.1002/ijgo.70326
13. Haight PJ, Riedinger CJ, Backes FJ, O'Malley DM, Cosgrove CM. The right time for change: a report on the heterogeneity of IVB endometrial cancer and improved risk-stratification provided by new 2023 FIGO staging criteria. Gynecol Oncol. 2023;175:32-40. https://doi.org/10.1016/j.ygyno.2023.05.069
14. Gravbrot N, Weil CR, DeCesaris CM, Gaffney DK, Suneja G, Burt LM. Differentiation of survival outcomes by anatomic involvement and histology with the revised 2023 International Federation of Gynecology and Obstetrics staging system for endometrial cancer. Eur J Cancer. 2024;201:113913. https://doi.org/10.1016/j.ejca.2024.113913
15. Matsuo K, Chen L, Klar M, Lee MW, Machida H, Mikami M, et al. Prognostic performance of the 2023 FIGO staging schema for endometrial cancer. Gynecol Oncol. 2024;187:37-45. https://doi.org/10.1016/j.ygyno.2024.04.016
16. Kobayashi-Kato M, Fujii E, Asami Y, Ahiko Y, Hiranuma K, Terao Y, et al. Utility of the revised FIGO2023 staging with molecular classification in endometrial cancer. Gynecol Oncol. 2023;178:178:36-43. https://doi.org/10.1016/j.ygyno.2023.09.011
17. Matsuo K, Klar M, Song BB, Roman LD, Wright JD. Validation of the 2023 FIGO staging schema for advanced endometrial cancer. Eur J Cancer. 2023;193:113316. https://doi.org/10.1016/j.ejca.2023.113316
18. Schwameis R, Fanfani F, Ebner C, Zimmermann N, Peters I, Nero C, et al. Verification of the prognostic precision of the new 2023 FIGO staging system in endometrial cancer patients - An international pooled analysis of three ESGO accredited centres. Eur J Cancer. 2023;193:113317. https://doi.org/10.1016/j.ejca.2023.113317

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251626

Evaluation of prognostic biomarkers for exacerbation frequency in patients with chronic obstructive pulmonary disease

Gulhan Ayhan Albayrak[1]*

SUMMARY

OBJECTIVE: Systemic inflammation is closely associated with acute exacerbations of chronic obstructive pulmonary disease and may reflect the risk of future exacerbations. Blood-based inflammatory biomarkers represent simple and practical tools for identifying patients at increased risk. The aim of this study was to evaluate the association between systemic inflammatory biomarkers measured during stable chronic obstructive pulmonary disease and overall exacerbation burden across Global Initiative for Chronic Obstructive Lung Disease stages.

METHODS: This retrospective cohort study included 132 patients with chronic obstructive pulmonary disease followed for at least 2 years. Clinical, functional, and laboratory data obtained during clinically stable periods were analyzed. Systemic inflammatory indices (neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio, systemic immune-inflammation index, systemic inflammation response index, aggregate index of systemic inflammation, and modified Glasgow prognostic score) were calculated. Exacerbation frequency was defined as the annualized number of moderate-to-severe exacerbations. Patients were classified according to Global Initiative for Chronic Obstructive Lung Disease groups A, B, and E, and associations between inflammatory biomarkers and clinical outcomes were evaluated using appropriate statistical methods, including receiver operating characteristic curve analyses.

RESULTS: Elevated levels of neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio, systemic immune-inflammation index, aggregate index of systemic inflammation, modified Glasgow prognostic score, and C-reactive protein were significantly associated with increased exacerbation frequency and hospitalization rates in the overall cohort (all $p<0.05$), whereas systemic inflammation response index showed no significant association. Receiver operating characteristic analyses demonstrated moderate discriminative ability of C-reactive protein and inflammatory indices for predicting frequent exacerbations.

CONCLUSION: Systemic inflammatory biomarkers were associated with adverse clinical outcomes in patients with chronic obstructive pulmonary disease and were significantly related to exacerbation frequency and severity. These findings suggest that systemic inflammatory markers may be useful for overall clinical assessment and identifying patients at increased risk of exacerbations, rather than for stage-specific risk stratification.

KEYWORDS: Chronic obstructive pulmonary disease. Inflammation. Exacerbations. Hospitalization. Prognosis. Diagnosis.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a chronic respiratory disease characterized by progressive airflow limitation in the lungs, and it's one of the leading causes of morbidity and mortality worldwide. Traditionally, the pathophysiology of COPD was thought to be limited to an inflammatory response against harmful particles or gases, such as those from chronic cigarette smoke exposure. However, recent studies have shown that this inflammatory process is not confined to the lungs but also affects systemic circulation, which in turn negatively impacts the course and prognosis of COPD[1,2].

Systemic inflammation is associated with important clinical outcomes of COPD, such as exacerbation frequency, disease severity, and comorbidities. This suggests that inflammation is not just a biological response but a fundamental mechanism that accelerates disease progression and severely reduces patients' quality of life. Therefore, identifying systemic inflammation, even in the stable phase of COPD, can play a significant role in predicting patients' future risk of exacerbation and overall survival[3].

Various systemic inflammation biomarkers, particularly ratios that are easily obtained from a complete blood count and provide information about the severity of inflammation, have recently gained significant interest in clinical practice. Some of these biomarkers include the neutrophil-to-lymphocyte Ratio (NLR), platelet-to-lymphocyte ratio (PLR), systemic immune-inflammation index (SII), systemic inflammation response index (SIRI), aggregate index of systemic inflammation (AISI), and modified Glasgow prognostic score (mGPS). It is known that these biomarkers have the potential to assess prognosis and disease severity in various chronic diseases. However, their prognostic value in COPD patients, especially regarding

[1]Health Sciences University, Şişli Hamidiye Etfal Training and Research Hospital, Department of Chest Diseases – Istanbul, Turkey.
*Corresponding author: gulhanayhanalbayrak@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on July 26, 2025. Accepted on January 10, 2026.
Scientific Editor: José Maria Soares Júnior

future exacerbation frequency and long-term survival, has not been fully elucidated[4].

This study aims to investigate the prognostic value of systemic inflammation biomarkers, measured during the stable phase, on disease severity, exacerbation frequency, and overall survival in patients diagnosed with COPD.

METHODS

Study design and population

A priori power analysis was performed before study initiation, and the required minimum sample size was calculated as 64 patients. This retrospective cohort study was conducted at a single tertiary training and research hospital and included 132 adult patients (≥18 years) with a confirmed diagnosis of COPD who met the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria and were followed in the outpatient clinic for at least 24 months. Blood samples and spirometry were obtained during routine visits when patients were clinically stable, defined as no acute exacerbation, hospitalization, or respiratory infection within the preceding 4 weeks.

Patients with conditions that could significantly affect systemic inflammatory markers (e.g., active infection, malignancy, hematologic, or chronic inflammatory diseases) or with incomplete laboratory or spirometric data were excluded. The study was approved by the institutional ethics committee and conducted in accordance with the Declaration of Helsinki.

Data collection and baseline assessments

Demographic and clinical data were retrospectively obtained from electronic medical records at study entry, including age, sex, smoking status, comorbidities, and COPD maintenance therapies (inhaled corticosteroids, long-acting β_2-agonists, and long-acting muscarinic antagonists). Pulmonary function test parameters—forced expiratory volume in one second (FEV_1), forced vital capacity (FVC), and the FEV_1/FVC ratio—were recorded from spirometry performed during routine outpatient visits in the clinically stable state closest to the time of blood sampling. Patients were classified according to Global Initiative for Chronic Obstructive Lung Disease (GOLD) stages based on post-bronchodilator FEV_1 values.

Laboratory measurements and inflammatory indices

Baseline laboratory data were obtained from venous blood samples collected during clinically stable outpatient visits and included complete blood count parameters (neutrophils, lymphocytes, monocytes, and platelets), C-reactive protein (CRP), and serum albumin levels. To minimize bias related to transient inflammatory conditions, medical records were reviewed to exclude concurrent infections or other acute illnesses at the time of sampling.

Systemic inflammatory indices were calculated using standard formulas

Systemic inflammatory indices were calculated using standard formulas: NLR (neutrophil count/lymphocyte count), PLR (platelet count/lymphocyte count), SII (platelet count×neutrophil count/lymphocyte count), SIRI (neutrophil count×monocyte count/lymphocyte count), AISI (neutrophil count×monocyte count×platelet count/lymphocyte count), and the mGPS. The mGPS was defined as follows: score 0, CRP≤10 mg/L and albumin ≥35 g/L; score 1, CRP>10 mg/L and albumin ≥35 g/L; and score 2, CRP>10 mg/L and albumin <35 g/L[5,6].

Clinical outcomes and follow-up

Clinical outcomes were evaluated over a fixed 24-month follow-up period after baseline assessment. The primary outcome was the annualized COPD exacerbation frequency, defined as the total number of moderate-to-severe exacerbations divided by the duration of follow-up in years. Moderate exacerbations required outpatient treatment with antibiotics and/or systemic corticosteroids, while severe exacerbations required hospitalization. Secondary outcomes included exacerbation-related hospitalization rates, total length of hospital stay, and all-cause mortality. In patients who died during follow-up, exacerbation rates were adjusted according to the available observation period to minimize underestimation.

Ethical declaration

Our institution has granted ethics committee approval with protocol number 2871/14.01.2025 As this was retrospective research, no informed consent was obtained from participants.

Statistical analysis

Statistical analyses were performed using International Business Machines Statistical Package for the Social Sciences Statistics for macOS, version 30.0 (IBM Corp., Armonk, NY). Categorical variables were summarized as frequencies and percentages, and continuous variables as mean±standard deviation or median (minimum–maximum), as appropriate. Normality was assessed using the Kolmogorov-Smirnov test. Intergroup comparisons were performed using the Kruskal-Wallis test for continuous variables and the chi-square or Fisher's exact test for categorical variables, with Bonferroni correction applied for post hoc analyses. Associations between continuous variables were evaluated

using Spearman's correlation analysis. Receiver operating characteristic (ROC) curve analyses were conducted to assess the predictive performance of systemic inflammatory biomarkers for frequent exacerbations, with area under the curve (AUC) values and optimal cut-off points calculated where appropriate. Statistical significance was defined as a two-sided $p<0.05$.

RESULTS

The demographic and clinical characteristics of the 132 patients stratified according to GOLD groups are presented in Table 1. The study population was predominantly elderly (mean age 72.0±7.0 years), male (75.8%), and ex-smokers (79.5%), with moderate-to-severe airflow limitation (mean FEV_1 43.2% predicted; FEV_1/FVC 61.8%). A history of prior exacerbations was present in 81.8% of patients, with a mean of 2.3 exacerbation events during follow-up.

Disease severity, symptom burden, exacerbation frequency, hospitalization rates, and systemic inflammatory indices increased progressively from GOLD Group A to Group E (all $p<0.001$), with Group E showing the most severe airflow limitation and the highest inflammatory marker levels. ROC analysis demonstrated that PLR, SII, and AISI significantly predicted exacerbations (all $p<0.05$), with PLR showing the best discriminative performance (AUC=0.691; cut-off >138.89, sensitivity 59.3%, specificity 79.2%). SII and AISI showed moderate discriminative ability for predicting exacerbations (AUC=0.636 and 0.630, respectively), with relatively high specificity. In contrast, NLR and SIRI were not statistically significant predictors; despite high sensitivity, NLR's low specificity limited its standalone utility (Table 2).

Correlation analyses (Table 3) demonstrated significant positive associations between exacerbation frequency and NLR, PLR, SII, and AISI, as well as between hospitalization rates and the same biomarkers (all $p<0.01$). SIRI was not significantly associated with exacerbation frequency and showed only a marginal association with hospitalizations. No significant correlations were observed within individual GOLD subgroups, likely reflecting limited sample size and event numbers.

DISCUSSION

This retrospective cohort study aimed to investigate the potential of systemic inflammatory markers—specifically NLR, PLR, SII, SIRI, AISI, and mGPS—to predict future exacerbation frequency, hospitalizations, and all-cause mortality in stable patients with COPD. Our findings demonstrate that NLR, PLR, SII, and AISI are significantly associated with a higher number of both exacerbations and hospitalizations in the overall study population. These results support the established understanding that systemic inflammation, a key component of COPD pathogenesis, plays a crucial role in determining the clinical progression and prognosis of the disease[7,8].

Numerous studies in the literature have examined the relationship between inflammatory markers and COPD exacerbations. For instance, some research has established that NLR and PLR are independent predictors of exacerbation frequency and mortality in COPD patients[9,10]. Our findings support and strengthen this evidence. Furthermore, our demonstration that next-generation inflammatory markers like SII and AISI are also significantly associated with exacerbations and hospitalizations is a crucial contribution. This finding supports the potential use of these simple and inexpensive tests as prognostic tools in routine clinical practice.

A notable finding in our study is that while systemic inflammatory markers showed a significant correlation with exacerbations and hospitalizations in the overall patient cohort, these correlations were not statistically significant within the individual GOLD subgroups (A, B, and E). This result suggests that the GOLD classification system, which primarily relies on airflow limitation and a patient's symptom and exacerbation history, may not fully capture or reflect the underlying inflammatory status of each subgroup. Consequently, this indicates that inflammatory markers might serve as a valuable supplement to the GOLD classification in a broader patient population. However, their predictive ability may be limited when applied to the specific, less severe characteristics of each distinct subgroup[11].

In a prospective study, the role of NLR and PLR in predicting frequent exacerbations in COPD was investigated. Blood parameters of patients in the stable phase were regularly monitored, and the values were compared when an exacerbation occurred. In patients prone to exacerbations, NLR and PLR values were found to be statistically significantly higher in the pre-exacerbation period compared to patients who did not experience an exacerbation. This suggests that these markers may reflect the pre-exacerbation inflammatory state[12].

A retrospective study examined the relationship between SII values and disease severity in stable COPD patients. It was found that SII values were significantly higher in patients with a more advanced GOLD stage compared to those with a lower stage. SII showed a positive correlation with traditional severity markers in COPD, such as the degree of airflow limitation (FEV1%) and symptom scores. This finding suggests that SII may reflect not only the risk of exacerbation but also the underlying severity of the disease[13].

Table 1. Distribution of patients' demographic and clinical findings according to Global Initiative for Chronic Obstructive Lung Disease stages.

Variables (n=132)	Gold-A (n=18)	Gold-B (n=45)	Gold-E (n=69)	p-value	Post-hoc*
	n (%) or median (min–max)	n (%) or median (min–max)	n (%) or median (min–max)		
Age (years)	70 (65–82)	68 (65–85)	71 (65–96)	0.179	
Sex n (%)				0.012	
Female	**2 (11.1)**	**6 (13.3)**	**24 (34.8)**		
Male	**16 (88.9)**	**39 (86.7)**	**45 (65.2)**		
BMI	26.7 (19–35.7)	26.1 (15.6–44.4)	25.5 (17.3–37.1)	0.474	
Smoking status				0.073	
Smoker	6 (33.3)	12 (26.7)	9 (13)		
Ex-smoker	12 (66.7)	33 (73.3)	60 (87)		
PFT findings					
FEV1 (%)	**60 (46–79)**	**49 (35–75)**	**34 (13–66)**	**<0.001**	**A–B; A–E; B–E**
FEV1/FVC (%)	72 (44–94)	69 (48–96)	72 (38–100)	0.665	
Exacerbation	**8 (44.4)**	**31 (68.9)**	**69 (100)**	**<0.001**	
Frequency of exacerbations	**0 (0–2)**	**1 (0–1)**	**3 (1–9)**	**<0.001**	**A–E; B–E**
Hospitalization	**0 (0)**	**0 (0)**	**41 (59.4)**	**<0.001**	
Number of hospitalizations	**0 (0–0)**	**0 (0–0)**	**1 (0–4)**	**<0.001**	**A–E; B–E**
Laboratory values					
Albumin (g/L)	**41.1 (34.6–45.9)**	**39.3 (29.4–53.6)**	**35.9 (27.9–44.4)**	**<0.001**	**A–E; B–E**
C-reactive protein (mg/L)	12.7 (0.9–55.7)	13.8 (0.8–32.6)	10.5 (1.1–72)	0.492	
White blood cell (×109/L)	7.4 (5.1–10.4)	8.3 (5–10.5)	8.3 (4.1–11)	0.316	
Hemoglobin (g/L)	13.2 (8.7–15.2)	13.7 (8.9–16.6)	13.2 (8.4–17)	0.328	
Hematocrit (%)	39.8 (26.7–45)	41.3 (27–50.2)	40.1 (26.5–49.6)	0.467	
Platelet (×109/L)	245.5 (108–485)	236 (108–337)	243 (117–821)	0.068	
Lymphocyte (×109/L)	1.8 (0.9–2.9)	1.7 (0.6–3.7)	1.7 (0.5–3.6)	0.459	
Neutrophil (×109/L)	4.7 (3.1–8.5)	5.2 (3–8.9)	5.8 (1.9–9.1)	0.125	
Monocyte (×109/L)	0.5 (0.2–1.2)	0.6 (0.1–3.8)	0.5 (0.3–1.4)	0.107	
Inflammation index					
NLR	2.7 (1.5–8.7)	2.7 (1.1–15.6)	3.7 (1.1–17.8)	0.065	
PLR	**137.4 (37.9–408.2)**	**131.3 (37.9–591.2)**	**163.8 (64.5–732.1)**	**0.015**	**B–E**
SII	**698.4 (184.5–3,453.8)**	**615 (184.5–5,273.8)**	**931.7 (221.3–5,930.4)**	**0.009**	**B–E**
SIRI	1.5 (0.6–4.5)	1.6 (0.5–14.8)	1.9 (0.3–10.5)	0.286	
AISI	**338 (126.5–1,796)**	**330.3 (123.1–4,166.3)**	**510.2 (76.1–3,321)**	**0.050**	**B–E**
mGPS				0.430	
0	7 (38.9)	13 (28.9)	32 (46.4)		
1	6 (33.3)	16 (35.6)	21 (30.4)		
2	5 (27.8)	16 (35.6)	16 (23.2)		

BMI: body mass index; NLR: neutrophil-to-lymphocyte ratio; PLR: platelet-to-lymphocyte ratio; SII: systemic immune-inflammation index; SIRI: systemic inflammation response index; AISI: aggregate index of systemic inflammation; mGPS: modified Glasgow prognostic score; PFT: Pulmonary Function Test(s); FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity. Bold values represent statistically significant differences ($p<0.05$). *indicates statistically significant results ($p<0.05$).

Table 2. Receiver operating characteristic analysis results of inflammatory biomarkers in predicting exacerbation development.

Variable	AUC (95%CI)	p-value	Cut-off	Sensitivity (%)	Specificity (%)
NLR	0.600 (0.469–0.730)	0.134	>1.9986	85.2	37.5
PLR	0.691 (0.528–0.791)	0.017	>138.89	59.3	79.2
SII	0.636 (0.510–0.763)	0.035	>810.16	49.1	79.2
SIRI	0.563 (0.431–0.694)	0.348	>1.41	69.4	54.2
AISI	0.630 (0.509–0.750)	0.035	>501.93	47.2	83.3

NLR: neutrophil-to-lymphocyte ratio; PLR: platelet-to-lymphocyte ratio; SII: systemic immune-inflammation index; SIRI: systemic inflammation response index; AISI: aggregate index of systemic inflammation; CI: confidence interval; AUC: area under the curve.

Table 3. Distribution of the relationship between exacerbation and hospitalization frequency and biomarkers.

Biomarkers		All patients		Gold-A		Gold-B		Gold-E	
		Exacerbation frequency	Hospital admissions	Exacerbation frequency	Hospital admissions	Exacerbation frequency	Hospital admissions	Exacerbation frequency	Hospital admissions
NLR	r	**0.218**	**0.248**	-0.136	–	0.174	–	0.065	0.199
	p	**0.012**	**0.004**	0.590	–	0.254	–	0.594	0.101
	n	**132**	**132**	18	18	45	45	69	69
PLR	r	**0.256**	**0.232**	0.102	–	0.277	–	0.025	0.155
	p	**0.003**	**0.008**	0.687	–	0.065	–	0.836	0.202
	n	**132**	**132**	18	18	45	45	69	69
SII	r	**0.278**	**0.273**	-0.115	–	0.203	–	0.071	0.176
	p	**0.001**	**0.002**	0.649	–	0.180	–	0.563	0.149
	n	**132**	**132**	18	18	45	45	69	69
SIRI	r	0.143	**0.171**	-0.195	–	0.118	–	0.097	0.157
	p	0.103	**0.050**	0.438	–	0.439	–	0.426	0.198
	n	132	**132**	18	18	45	45	69	69
AISI	r	**0.242**	**0.230**	-0.055	–	0.170	–	0.098	0.156
	p	**0.005**	**0.008**	0.828	–	0.264	–	0.425	0.201
	n	**132**	**132**	18	18	45	45	69	69
mGPS	r	-0.138	-0.081	0.168	–	-0.130	–	0.003	0.037
	p	0.116	0.356	0.505	–	0.396	–	0.980	0.764
	n	132	132	18	18	45	45	69	69

NLR: neutrophil-to-lymphocyte ratio; PLR: platelet-to-lymphocyte ratio; SII: systemic immune-inflammation index; SIRI: systemic inflammation response index; AISI: aggregate index of systemic inflammation; mGPS: modified Glasgow prognostic score. Bold values represent statistically significant differences ($p<0.05$).

A study showed that a high mGPS score is associated with poor clinical outcomes (the need for intensive care unit transfer and in-hospital mortality) in patients with Acute Exacerbation of Chronic Obstructive Pulmonary Disease (AE-COPD). This study investigated the value of mGPS in predicting mortality in patients hospitalized for acute COPD exacerbation. Specifically, patients with a GPS score of 2 were found to have a higher risk of adverse outcomes compared to those with lower scores. This suggests that mGPS may be an independent risk factor for mortality in AE-COPD[14].

One of the main strengths of our study is that it examined not only classic markers (NLR and PLR) but also newer and more comprehensive markers like SII and AISI. Additionally, the detailed examination of the followed patient group and clinical outcomes increases the reliability of our findings. Several contemporary studies have reported associations between blood-based inflammatory markers—particularly NLR and PLR—and exacerbation frequency, hospitalization, and mortality in COPD[15]. Composite indices such as SII and AISI, which integrate multiple immune cell populations, have recently gained

attention as more comprehensive markers of systemic inflammation and immune dysregulation[16]. The observed associations in our study support the biological plausibility that these indices may capture the cumulative inflammatory burden more effectively than single biomarkers.

Limitations

This study has several limitations. Its retrospective, single-center design limits causal inference and generalizability. Although blood samples were obtained during clinically stable periods and patients with recent exacerbations or infections were excluded, subclinical inflammation at the time of sampling cannot be fully ruled out. Residual confounding from unmeasured inflammatory conditions may also have influenced biomarker levels. In addition, no a priori power analysis was performed, and subgroup analyses may have been underpowered. Detailed longitudinal data on treatment changes, adherence, and the use of add-on therapies were incomplete. Finally, although mortality was considered in exacerbation rate calculations, competing risks related to death may still have affected the results. Despite these limitations, the study provides clinically relevant real-world data on readily available inflammatory biomarkers in COPD.

CONCLUSION

Systemic inflammatory biomarkers measured during stable COPD were associated with future exacerbation burden, with higher baseline CRP and composite indices linked to more frequent moderate-to-severe exacerbations. Although causality cannot be inferred due to the retrospective design, these readily available biomarkers may serve as practical adjuncts for risk stratification in routine clinical practice. Prospective multicenter studies are needed to validate these findings and clarify their role in personalized COPD management.

CONSENT FOR PUBLICATION

The original article is not under consideration by another publication, and its substance, tables, or figures have not been published previously and will only be published elsewhere.

ETHICAL DECLARATION

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008. Our institution has granted ethics committee approval with protocol number 2871/14.01.2025. As this was retrospective research, no informed consent was obtained from participants.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Erhabor GE, Adeniyi B, Arawomo AO, Akinwalere O, Adetona G, Fagbohun FT, et al. Acute exacerbation of COPD: clinical perspectives and literature review. West Afr J Med. 2021;38(11):1129-42. PMID: 34922414
2. MacLeod M, Papi A, Contoli M, Beghé B, Celli BR, Wedzicha JA, et al. Chronic obstructive pulmonary disease exacerbation fundamentals: diagnosis, treatment, prevention and disease impact. Respirology. 2021;26(6):532-51. https://doi.org/10.1111/resp.14041
3. Gogol M, Krawczyk J. Systemic inflammation and its impact on the prognosis of patients with chronic obstructive pulmonary disease. J Clin Med. 2022;11(18):5366. https://doi.org/10.3390/jcm11185366
4. López-Campos JL, Cosío BG. The systemic inflammatory response in COPD and its clinical implications. Arch Bronconeumol. 2021;57(11):743-5. https://doi.org/10.1016/j.arbres.2021.04.006
5. Fu Y, Wang Y, Wang Y, Mou T, He X, Wang J, et al. Biomarkers (NLR, PLR, SII) for frequent COPD exacerbations: diagnostic and clinical management implications in a retrospective study. Int J Chron Obstruct Pulmon Dis. 2025;20:987-98. https://doi.org/10.2147/COPD.S510118
6. Karataş H, Koçak C, Çiftçi B. Evaluation of the systemic inflammatory response index in patients with chronic obstructive pulmonary disease. J Thorac Dis. 2022;14(4):1675-83. https://doi.org/10.21037/jtd-21-1789
7. Luo Z, Zhang W, Chen L, Xu N. Prognostic value of neutrophil:lymphocyte and platelet:lymphocyte ratios for 28-day mortality of patients with AECOPD. Int J Gen Med. 2021;14:2839-48. https://doi.org/10.2147/IJGM.S312045
8. Yao C, Liu X, Tang Z. Prognostic role of neutrophil-lymphocyte ratio and platelet-lymphocyte ratio for hospital mortality in patients with AECOPD. Int J Chron Obstruct Pulmon Dis. 2017;12:2285-90. https://doi.org/10.2147/COPD.S141760
9. Yang YL, Wu CH, Hsu PF, Lee TM, Chen SC, Lee CS. et al. Systemic immune-inflammation index as a predictor of clinical outcomes in chronic inflammatory conditions. Respir Med. 2022;191:106709. https://doi.org/10.1016/j.rmed.2021.106709
10. Cai C, Zeng W, Wang H, Ren S. Neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR) and monocyte-to-lymphocyte ratio (MLR) as biomarkers in diagnosis evaluation of acute exacerbation of chronic obstructive pulmonary disease: a retrospective, observational study. Int J Chron Obstruct Pulmon Dis. 2024;19:933-43. https://doi.org/10.2147/COPD.S452444

11. Agustí A, Celli BR, Criner GJ, Halpin D, Anzueto A, Barnes P, et al. Global initiative for chronic obstructive lung disease 2023 report: GOLD executive summary. Eur Respir J. 2023;61(4):2300239. https://doi.org/10.1183/13993003.00239-2023
12. Liao QQ, Mo YJ, Zhu KW, Gao F, Huang B, Chen P, et al. Platelet-to-lymphocyte ratio (PLR), neutrophil-to-lymphocyte ratio (NLR), monocyte-to-lymphocyte ratio (MLR), and eosinophil-to-lymphocyte ratio (ELR) as biomarkers in patients with acute exacerbation of chronic obstructive pulmonary disease (AECOPD). Int J Chron Obstruct Pulmon Dis. 2024;19:501-18. https://doi.org/10.2147/COPD.S447519
13. Li D, Liu L, Lv J, Xiong X. The clinical significance of peripheral blood-related inflammatory markers in patients with AECOPD. Immunobiology. 2025;230(3):152903. https://doi.org/10.1016/j.imbio.2025.152903
14. Kuluöztürk M, Deveci F. The Glasgow prognostic score can be a predictor of mortality in acute exacerbation of chronic obstructive pulmonary disease. Expert Rev Respir Med. 2020;14(5):521-5. https://doi.org/10.1080/17476348.2020.1735366
15. Paliogiannis P, Fois AG, Sotgia S, Mangoni AA, Zinellu E, Pirina P, et al. Neutrophil-to-lymphocyte ratio and clinical outcomes in COPD: recent evidence and future perspectives. Clin Chem Lab Med. 2020;58(12):1901-9. https://doi.org/10.1515/cclm-2019-1173021.106709
16. Hu B, Yang XR, Xu Y, Sun YF, Sun C, Guo W, et al. Systemic immune-inflammation index predicts prognosis in inflammatory diseases. J Inflamm Res. 2021;14:303-15. https://doi.org/10.2147/JIR.S295789

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251628

Elevated serum Raftlin levels in non-muscle invasive bladder cancer: diagnostic implications and rs690037 genotype analysis

Senay Balci[1]*, Rojda Tanriverdi[1], Ali Nebioğlu[2], Didem Derici Yildirim[3], Cemil Gülüm[4], Murat Bozlu[5], Lülüfer Tamer[1]

SUMMARY

OBJECTIVE: Non-muscle invasive bladder cancer is a common urological malignancy characterized by high recurrence and progression rates. There is a critical need for non-invasive biomarkers to support early detection and disease monitoring. Raftlin, a lipid raft-associated scaffold protein involved in immune regulation and signal transduction, has recently gained attention in cancer biology. The aim of this study was to investigate serum Raftlin levels and the rs690037 polymorphism of the Raftlin gene in non-muscle invasive bladder cancer patients.

METHODS: A case-control study was conducted, including 30 patients diagnosed with non-muscle invasive bladder cancer and 50 healthy controls. Serum Raftlin levels were measured using enzyme-linked immunosorbent assay, and the rs690037 polymorphism was analyzed using real-time polymerase chain reaction . Gender-stratified analysis and receiver operating characteristic curve analysis were performed. Clinical and pathological characteristics including tumor stage, grade, and European Organisation for Research and Treatment of Cancer risk groups, were analyzed.

RESULTS: Serum Raftlin levels were significantly elevated in non-muscle invasive bladder cancer patients compared to healthy controls ($p<0.001$). Gender-stratified analysis confirmed this association was independent of gender (males: $p<0.001$; females: $p<0.001$). Receiver operating characteristic curve analysis demonstrated an area under the curve of 0.987 (95%CI 0.966–1.000), with an optimal cut-off of 5.28 ng/mL providing 96.7% sensitivity and 98.0% specificity. Trends toward higher Raftlin levels were observed in more aggressive disease features. No significant association was found between the rs690037 polymorphism and disease presence.

CONCLUSION: This pilot study suggests that serum Raftlin levels are significantly elevated in non-muscle invasive bladder cancer and show high discriminatory ability. While preliminary findings suggest potential as a non-invasive biomarker, validation in larger, independent cohorts is required before clinical implementation.

KEYWORDS: Raftlin. Non-muscle invasive bladder cancer. Biomarker. RFTN1. Polymorphism.

INTRODUCTION

Bladder cancer (BC) is one of the most common malignancies worldwide, with an incidence approximately three times higher in males than in females[1,2]. Non-muscle invasive bladder cancer (NMIBC), encompassing tumors limited to the mucosa or lamina propria, represents about 75% of all BC cases[3]. NMIBC poses substantial clinical challenges due to its high recurrence rate (50–70%) and risk of progression to muscle-invasive disease (10–30%)[2,4,5]. Therefore, there is an urgent need to develop non-invasive biomarkers for early diagnosis and prediction of disease progression[4].

In recent years, protein-based biomarkers have gained considerable attention in the diagnosis and monitoring of BC[6]. Among these, lipid raft-associated proteins have emerged as promising candidates. Lipid rafts are specialized membrane microdomains that facilitate cell signaling, receptor clustering, and intercellular communication. These microdomains are linked to several oncogenic pathways such as Insulin-like Growth Factor, Phosphoinositide 3-kinase/Protein kinase B, Vascular Endothelial Growth Factor/Vascular Endothelial Growth Factor Receptor 2, and Cluster of Differentiation 44, and are known to regulate tumor proliferation, migration, and drug resistance[7,8].

Additionally, lipid rafts represent critical regulatory platforms in the context of cancer stem cells (CSCs) within the tumor microenvironment. Evidence shows that CSCs exhibit

[1]Mersin University, Faculty of Medicine, Department of Biochemistry – Mersin, Turkey.
[2]Mersin City Training and Research Hospital, Department of Urology – Mersin, Turkey.
[3]Mersin University, Faculty of Medicine, Department of Biostatistics and Medical Informatics – Mersin, Turkey.
[4]Mersin University, Department of Chemistry – Mersin, Turkey.
[5]Mersin University, Department of Urology – Mersin, Turkey.
*Corresponding author: sbfidanci@hotmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: this project was supported by the Health Institutes of Turkey (TÜSEB), call number 2024-A4-01, project number 39800.
Received on January 29, 2026. Accepted on March 03, 2026.
Scientific Editor: José Maria Soares Júnior

significantly higher lipid raft levels compared to other cancer cells, and CSC markers such as CD24, CD44, and CD133 are localized to these microdomains. Lipid rafts facilitate CSC self-renewal, epithelial–mesenchymal transition (EMT), drug resistance, and niche maintenance[9].

Raftlin (RFTN1), a scaffold protein for lipid rafts, plays a key role in B-cell receptor (BCR) signaling[8]. Recent insights suggest that RFTN1 may also modulate inflammatory mechanisms mediated by Toll-like receptor (TLR) pathways and immune cell differentiation. Specifically, RFTN1 interacts with nuclear receptor Retinoic acid receptor-related orphan receptor gamma t (RORγt) to promote interleukin 17 (IL-17) expression and enhance pathogenic Th17 differentiation, linking it to pro-inflammatory processes[9]. By supporting lipid raft integrity, RFTN1 may influence inflammation-driven tumor microenvironments.

Emerging studies also implicate RFTN1 in tumor cell proliferation, migration, and cellular adhesion. For instance, RFTN1 overexpression has been associated with increased proliferation in non-small cell lung cancer (NSCLC)[10], and involvement in adhesion-related signaling in various cancers[11].

Although direct data on the RFTN1 rs690037 polymorphism are limited, genetic variants such as this may affect protein expression or function, potentially influencing cancer susceptibility. The rs690037 variant has been previously associated with primary open-angle glaucoma[12], suggesting potential functional consequences that warrant investigation in other disease contexts.

This study aimed to evaluate serum RFTN1 levels in NMIBC patients and investigate the association of the rs690037 polymorphism with disease risk. Moreover, we analyzed the relationship between these parameters and clinical/pathological features to assess RFTN1's potential as a biomarker in NMIBC.

METHODS

Study population

Between February 2024 and June 2024, data from 30 patients who were admitted to the Urology Clinic of Mersin University and followed and treated with a diagnosis of NMIBC, as well as 50 healthy individuals with no history of malignancy, were prospectively analyzed. Participants with invasive and/or metastatic BC, those under the age of 18, patients with malignancies other than BC, and individuals who did not provide informed consent were excluded from the study.

Sociodemographic data, including age, weight, height, body mass index (BMI), family history, and smoking status, were collected for all participants included in the study. Written informed consent forms were obtained from all individuals participating in the study.

Specific ethical approval, numbered 2024/089 and dated January 31, 2024, was obtained from the Clinical Research Ethics Committee (Institutional Review Board) of Mersin University Faculty of Medicine, focusing on RFTN1 analysis in bladder cancer, in accordance with the Helsinki Declaration for the current study.

Sample collection and storage

The analysis of serum RFTN1 levels and the detection of the RFTN1 gene polymorphism rs690037 were performed at the Medical Biochemistry Laboratory of Mersin University. For polymorphism analysis, blood samples were collected from both patients and healthy individuals into 2 mL ethylenediaminetetraacetic acid (EDTA) tubes. For the enzyme-linked immunosorbent assay (ELISA), blood samples were drawn into 5 mL plain biochemistry tubes.

The blood samples in the plain biochemistry tubes were centrifuged at 4,000 rpm for 10 min, and the serum was separated. Storage of serum samples at -80°C retarded protein degradation, thereby preserving the long-term stability of biomolecules such as RFTN1 and enhancing the reliability of the research results. Therefore, serum samples were maintained at -80°C until the day of analysis. Blood samples collected in EDTA tubes were stored at +4°C.

Measurement of Raftlin levels

Serum RFTN1 levels were measured using a commercial BT Lab ELISA kit (E4717Hu). Assay precision was evaluated by repeated measurements of random samples within the same batch, and intra-assay precision was assessed by analyzing three serum samples (n=18 each), yielding coefficients of variation (CV%) of 6.1, 5.4, and 5.9%, respectively.

The assay sensitivity was 0.41 ng/mL, with a detection range of 0.5–40 ng/mL. Standard curves were generated using recombinant RFTN1 standards at concentrations of 0, 1.5, 3, 6, 12, and 24 ng/mL, in accordance with the manufacturer's protocol.

This process ensures that the quantification of RFTN1 is both accurate and reliable, as it is based on comparison with established standards and validated analytical performance. The utilization of a four-parameter logistic model represents the standard approach in ELISA assays to translate optical density readings into precise concentration values across the full dynamic range of the assay.

DNA isolation and polymorphism analysis

DNA isolation from blood samples stored at +4°C was performed using a commercial DNA isolation kit (Roche Diagnostics,

Mannheim, Germany). The isolated DNA samples were then analyzed for the RFTN1 gene polymorphism rs690037 using a real-time polymerase chain reaction (PCR) device (Roche LightCycler 480; Roche Diagnostics, Mannheim, Germany).

This approach ensures high-quality DNA extraction and precise detection of genetic polymorphisms. Real-time PCR represents a sensitive and specific method widely employed for genotyping, enabling accurate identification of the rs690037 polymorphism in the RFTN1 gene.

Statistical analysis

Statistical analyses of the data obtained in this study were performed using the trial version of the Statistical Package for the Social Sciences (SPSS) 25.0 software package (SPSS, IBM Corp., Armonk, NY, USA). The distribution of continuous variables was assessed with the Kolmogorov-Smirnov test. For data that did not show a normal distribution, descriptive statistics were presented as median and interquartile range (25th and 75th percentiles). Categorical variables were expressed as numbers and percentages (%).

To compare gender and genotype distributions between groups, the chi-square test was applied. For continuous variables such as age and serum RFTN1 levels, the non-parametric Mann-Whitney U test was used. Gender-stratified analysis was performed to evaluate RFTN1 levels separately in male and female subgroups. Hardy-Weinberg equilibrium analysis for genetic data was evaluated separately for patient and control groups using the chi-square test.

Receiver operating characteristic (ROC) curve analysis was performed to assess the diagnostic performance of serum RFTN1 levels in discriminating NMIBC patients from healthy controls. The area under the curve (AUC), optimal cut-off value, sensitivity, and specificity were calculated. The optimal cut-off point was determined using the Youden Index (sensitivity+specificity-1). In all statistical analyses, a p-value of less than 0.05 was considered statistically significant.

RESULTS

A total of 80 individuals were included in the study. Of these, 30 comprised the patient group diagnosed with NMIBC, while 50 constituted the healthy control group. The distribution of participants' age, gender, sociodemographic, and clinical characteristics is summarized in Table 1. There was no statistically significant difference in age between the patient and control groups (p=0.313). However, the proportion of males was found to be significantly higher in the patient group compared to the control group (p=0.001).

Regarding genotype distributions, no statistically significant difference was observed between the groups (p=0.439). In addition, Hardy-Weinberg equilibrium analysis revealed no significant deviation in either group (p=0.284 for the patient group, p=0.428 for the control group) (Table 2).

Serum RFTN1 levels did not exhibit a normal distribution in either the patient or control groups (p<0.001). Therefore, descriptive statistics are presented as median and 25th–75th percentiles. Serum RFTN1 levels were found to be statistically significantly elevated in the NMIBC group compared to the control group (p<0.001). The distribution of serum RFTN1 levels according to groups is presented in Table 3.

Given the significant gender imbalance between groups (76.7% male in NMIBC vs. 38.0% in controls), a gender-stratified analysis was performed. In male subjects, serum RFTN1 levels were significantly higher in NMIBC

Table 2. Distribution of RS690037 genotypes in patient and control groups.

Genotype	Control group n (%)	NMIBC group n (%)	Total n (%)	p-value
TT	12 (24.0%)	5 (16.7%)	17 (21.3%)	0.439
TC	21 (42.0%)	17 (56.7%)	38 (47.5%)	
CC	17 (34.0%)	8 (26.7%)	25 (31.3%)	

NMIBC: non-muscle invasive bladder cancer; TT: homozygous T allele; TC: heterozygous (T/C); CC: homozygous C allele.

Table 1. Sociodemographic and clinical characteristics of the study groups.

Variable	Control group (n=50)	NMIBC group (n=30)	Total (n=80)	p-value
Gender, n (%)				
Female	31 (62.0%)	7 (23.3%)	38 (47.5%)	**0.001**
Male	19 (38.0%)	23 (76.7%)	42 (52.5%)	
Age, mean±SD (years)	65.44±5.54	66.87±6.92	–	0.313
Comorbid disease, n (%)	20 (40.0%)	12 (40.0%)	32 (40.0%)	0.999
BMI, mean±SD (kg/m²)	26.2±2.1	26.1±2.0	–	0.872
Family history, n (%)	15 (30.0%)	9 (30.0%)	24 (30.0%)	0.999
Smoking, n (%)	25 (50.0%)	15 (50.0%)	40 (50.0%)	0.999

NMIBC: non-muscle invasive bladder cancer; BMI: body mass index; SD: standard deviation. Bold values indicate statistically significant differences (p<0.05).

patients (median: 7.48 ng/mL, interquartile range [IQR]: 6.92–8.45) compared to male controls (median: 3.15 ng/mL, IQR: 2.40–3.52; p<0.001). Similarly, in female subjects, RFTN1 levels were significantly elevated in NMIBC patients (median: 8.12 ng/mL, IQR: 7.35–9.20) compared to female controls (median: 3.24 ng/mL, IQR: 2.51–3.64; p<0.001). These results are summarized in Table 4, demonstrating that the association between RFTN1 and NMIBC is independent of gender.

Within the NMIBC patient group (n=30), clinical and pathological data were available and analyzed. The distribution included 18 patients (60.0%) with Ta stage tumors, 10 patients (33.3%) with T1 stage, and 2 patients (6.7%) with carcinoma in situ (CIS). Regarding tumor grade, 11 patients (36.7%) had low-grade tumors and 19 patients (63.3%) had high-grade tumors. According to European Organisation for Research and Treatment of Cancer (EORTC) risk stratification, 9 patients (30.0%) were classified as low risk, 12 patients (40.0%) as intermediate risk, and 9 patients (30.0%) as high risk.

Serum RFTN1 levels showed a trend toward higher values in patients with more aggressive disease features, though statistical significance was limited by sample size. RFTN1 levels were numerically higher in T1 tumors (median: 8.05 ng/mL) compared to Ta tumors (median: 7.38 ng/mL), and in high-grade tumors (median: 7.89 ng/mL) compared to low-grade tumors (median: 7.21 ng/mL). When stratified by EORTC risk groups, RFTN1 levels showed progressive elevation from low-risk (median: 7.15 ng/mL) to intermediate-risk (median: 7.62 ng/mL) to high-risk patients (median: 8.28 ng/mL). These findings are presented in Table 5.

Table 3. Serum Raftlin levels and tests of normality in patient and control groups.

Group	n	Median (ng/mL)	25th percentile	75th percentile	Minimum	Maximum	p-value
Control	50	3.20	2.47	3.59	1.26	4.52	<0.001
NMIBC	30	7.62	7.05	8.61	5.04	12.20	

NMIBC: non-muscle invasive bladder cancer. Note: Both Kolmogorov-Smirnov and Shapiro-Wilk tests indicate that serum RFTN1 levels do not follow a normal distribution (p<0.05). Therefore, serum RFTN1 levels are presented as median [25th–75th percentile]. The Mann-Whitney U test was used for comparison.

Table 4. Gender-stratified analysis of serum Raftlin levels.

Gender	Group	N	Median (ng/mL)	25th percentile	75th percentile	p-value
Male	Control	19	3.15	2.40	3.52	<0.001
	NMIBC	23	7.48	6.92	8.45	
Female	Control	31	3.24	2.51	3.64	<0.001
	NMIBC	7	8.12	7.35	9.20	

NMIBC: non-muscle invasive bladder cancer. Note: The Mann-Whitney U test was used for comparison within each gender subgroup.

Table 5. Clinical and pathological characteristics of non-muscle invasive bladder cancer patients and association with serum Raftlin levels.

Characteristic	n (%)	Median RFTN1 (ng/mL)	25th percentile	75th percentile
Tumor stage				
Ta	18 (60.0%)	7.38	6.85	8.42
T1	10 (33.3%)	8.05	7.28	8.95
CIS	2 (6.7%)	7.82	7.15	8.49
Tumor grade				
Low-grade	11 (36.7%)	7.21	6.78	7.95
High-grade	19 (63.3%)	7.89	7.18	8.82
EORTC risk group				
Low risk	9 (30.0%)	7.15	6.68	7.88
Intermediate risk	12 (40.0%)	7.62	7.05	8.45
High risk	9 (30.0%)	8.28	7.52	9.28

CIS: carcinoma in situ; EORTC: European Organisation for Research and Treatment of Cancer; RFTN1: Raftlin.

ROC curve analysis was performed to evaluate the diagnostic performance of serum RFTN1 in discriminating NMIBC patients from healthy controls. The analysis yielded an AUC of 0.987 (95%CI 0.966–1.000; p<0.001), indicating a high level of discriminatory performance within this study population. Using the Youden Index, the optimal cut-off value was determined to be 5.28 ng/mL, which provided a sensitivity of 96.7% (95%CI 82.8–99.9%) and a specificity of 98.0% (95%CI 89.4–99.9%) for detecting NMIBC. The positive predictive value was 96.7%, and the negative predictive value was 98.0%. These results are summarized in Figure 1.

In summary, serum RFTN1 levels were found to be significantly elevated in the NMIBC patient group across both genders. The gender-stratified analysis confirmed that elevated RFTN1 is independently associated with NMIBC. Clinical correlation analysis revealed trends toward higher RFTN1 levels in more aggressive disease subtypes. The ROC curve analysis demonstrated excellent diagnostic performance, supporting the potential utility of serum RFTN1 as a biomarker for NMIBC detection.

DISCUSSION

In this study, the relationship between serum RFTN1 levels and the rs690037 polymorphism of the RFTN1 gene was evaluated in patients diagnosed with NMIBC. The results demonstrated a statistically significant elevation in serum RFTN1 levels in the NMIBC group compared to healthy controls, while no significant differences were observed in the distribution of rs690037 genotypes. These findings highlight the potential diagnostic relevance of RFTN1 as a non-invasive biomarker in BC and warrant further mechanistic investigations.

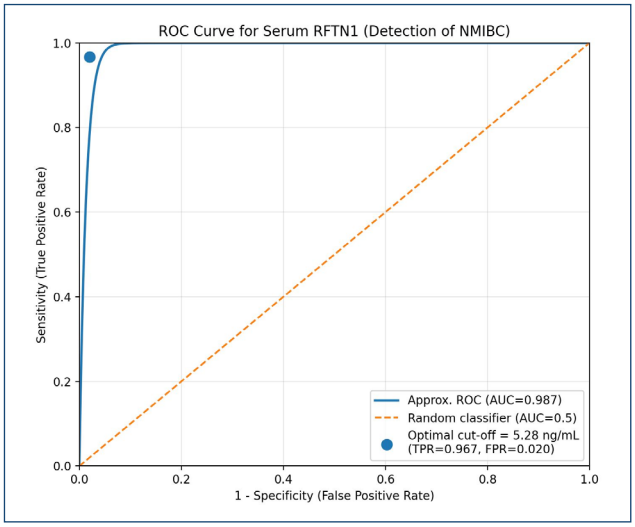


Figure 1. Diagnostic performance of serum Raftlin for detection of non-muscle invasive bladder cancer. AUC: area under the curve. Note: Optimal cut-off value was determined using the Youden Index.

RFTN1 is a scaffold protein localized within lipid rafts—specialized microdomains of the plasma membrane involved in protein clustering, receptor organization, and signal transduction. Originally characterized for its role in BCR signaling[8], RFTN1 has more recently emerged as a modulator of signaling cascades relevant to cancer biology. Elevated RFTN1 expression has been identified in several solid tumors, including NSCLC[10] and other malignancies[11]. The increased RFTN1 levels observed in NMIBC patients may reflect tumor-driven upregulation or active release into the bloodstream.

RFTN1 has been implicated in immunomodulatory pathways. Recent evidence indicates that RFTN1 interacts with nuclear receptor RORγt to promote IL-17 expression and enhance pathogenic Th17 differentiation[9], linking it to pro-inflammatory processes in the tumor microenvironment[13]. RFTN1 has also been implicated in angiogenesis, where it is recruited to the VEGFR2 complex via neuropilin-1 (NRP1), facilitating VEGF-mediated angiogenic signaling[14] a mechanism relevant to tumor growth and vascular remodeling.

The gender-stratified analysis in our study demonstrated that elevated RFTN1 levels were significantly associated with NMIBC in both male and female subgroups, confirming that this association is independent of the observed gender imbalance between groups. Furthermore, the ROC curve analysis yielded an AUC of 0.987, indicating a high level of discriminatory performance for distinguishing NMIBC patients from controls in the present study. The optimal cut-off of 5.28 ng/mL provided sensitivity of 96.7% and specificity of 98.0%, supporting the potential diagnostic utility of serum RFTN1.

Clinical correlation analysis revealed trends toward higher RFTN1 levels in patients with more aggressive disease features, including T1 stage, high-grade tumors, and high EORTC risk classification. However, statistical significance was limited by the modest sample size, and these observations require validation in larger cohorts.

The absence of a statistically significant difference in rs690037 genotype distributions between patients and controls suggests that the observed increase in serum RFTN1 levels may not be primarily driven by this specific genetic variant. RFTN1 regulation in NMIBC is more likely mediated through epigenetic modifications, post-translational alterations, or tumor microenvironmental stimuli. The rs690037 variant, although previously associated with primary open-angle glaucoma[12], may lack functional consequences in the context of BC, or the sample size may have been insufficient to detect small genetic effects.

Our findings support the hypothesis that RFTN1 is involved in the pathophysiology of NMIBC. The clinical utility of RFTN1 as a biomarker may be enhanced by integrating it into a panel with other novel markers such as RNF19A[15]. Future studies should evaluate RFTN1 levels alongside tumor stage, grade, recurrence risk, and treatment response in larger, prospective cohorts.

CONCLUSION

This pilot study suggests that serum RFTN1 levels are significantly elevated in patients with NMIBC compared to healthy controls, and this association remained significant after gender-stratified analysis. ROC curve analysis indicated a high level of discriminatory performance (AUC: 0.987), with high sensitivity and specificity in the present study. Clinical correlation analysis revealed trends toward higher RFTN1 levels in more aggressive disease features; however, these observations were limited by the small sample size. The rs690037 polymorphism showed no significant association with disease susceptibility, suggesting that RFTN1 regulation in NMIBC may be influenced predominantly by non-genetic mechanisms.

Despite the promising preliminary findings, several important limitations must be acknowledged. The modest sample size (n=30 patients) limits statistical power, particularly for subgroup analyses and genetic associations. The gender imbalance between groups, while addressed through stratified analysis, highlights the need for matched control groups in future studies. Additionally, the cross-sectional design precludes assessment of RFTN1's utility in monitoring disease recurrence or progression.

Future validation studies in larger, independent cohorts are essential to confirm these findings and establish RFTN1's clinical utility. Prospective studies should evaluate RFTN1 levels in relation to disease recurrence, progression, and treatment response. Integration of RFTN1 into multi-marker panels, potentially including established biomarkers and emerging candidates such as RNF19A[15], may enhance diagnostic and prognostic accuracy. Functional studies are also warranted to elucidate the mechanistic role of RFTN1 in NMIBC pathophysiology. Only after such comprehensive validation can RFTN1 be considered for clinical implementation as a biomarker in NMIBC management.

ETHICAL APPROVAL

Approved by the Clinical Research Ethics Committee of Mersin University Faculty of Medicine (Decision date: 31.01.2024; Decision number: 2024/089).

AUTHORS' CONTRIBUTIONS

SB: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. **RT:** Data curation, Investigation, Resources. **AN:** Investigation, Writing – review & editing. **DDY:** Formal Analysis, Visualization. **CG:** Data curation, Investigation, Software. **MB:** Supervision, Writing – review & editing. **LT:** Methodology, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. 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(3):209-49. https://doi.org/10.3322/caac.21660
2. Burger M, Catto JW, Dalbagni G, Grossman HB, Herr H, Karakiewicz P, et al. Epidemiology and risk factors of urothelial bladder cancer. Eur Urol. 2013;63(2):234-41. https://doi.org/10.1016/j.eururo.2012.07.033
3. Babjuk M, Burger M, Compérat EM, Gontero P, Mostafid AH, Palou J, et al. European Association of Urology Guidelines on non-muscle-invasive bladder cancer (TaT1 and Carcinoma In Situ) - 2019 update. Eur Urol. 2019;76(5):639-57. https://doi.org/10.1016/j.eururo.2019.08.016
4. Sylvester RJ, Meijden AP, Oosterlinck W, Witjes JA, Bouffioux C, Denis L, et al. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. Eur Urol. 2006;49(3):466-5;discussion 475-7. https://doi.org/10.1016/j.eururo.2005.12.031
5. Cambier S, Sylvester RJ, Collette L, Gontero P, Brausi MA, Andel G, et al. EORTC Nomograms and Risk Groups for predicting recurrence, progression, and disease-specific and overall survival in non-muscle-invasive stage Ta-T1 urothelial bladder cancer patients treated with 1-3 years of maintenance Bacillus Calmette-Guérin. Eur Urol. 2016;69(1):60-9. https://doi.org/10.1016/j.eururo.2015.06.045
6. Schmitz-Dräger BJ, Droller M, Lokeshwar VB, Lotan Y, Hudson MA, Rhijn BW, et al. Molecular markers for bladder cancer screening, early diagnosis, and surveillance: the WHO/ICUD consensus. Urol Int. 2015;94(1):1-24. https://doi.org/10.1159/000369357

7. Li B, Qin Y, Yu X, Xu X, Yu W. Lipid raft involvement in signal transduction in cancer cell survival, cell death and metastasis. Cell Prolif. 2022;55(1):e13167. https://doi.org/10.1111/cpr.13167
8. Saeki K, Miura Y, Aki D, Kurosaki T, Yoshimura A. The B cell-specific major raft protein, Raftlin, is necessary for the integrity of lipid raft and BCR signal transduction. EMBO J. 2003;22(12):3015-26. https://doi.org/10.1093/emboj/cdg293
9. Singh AK, Kumar R, Yin J, Brooks JF II, Kathania M, Mukherjee S, et al. RORγt-Raftlin1 complex regulates the pathogenicity of Th17 cells and colonic inflammation. Nat Commun. 2023;14(1):4972. https://doi.org/10.1038/s41467-023-40622-1
10. Eliçora A, Güven B, Engin H, Çevik GT, Sezer HF. Could serum Raftlin and GPER-1 levels be new biomarkers for early detection of non-small cell lung cancer? J Cardiothorac Surg. 2025;20(1):92. https://doi.org/10.1186/s13019-024-03296-4
11. Zhao Y, Gao Y, Xu X, Zhou J, Wang H. Multi-omics analysis of genomics, epigenomics and transcriptomics for molecular subtypes and core genes for lung adenocarcinoma. BMC Cancer. 2021;21(1):257. https://doi.org/10.1186/s12885-021-07888-4
12. Chen JH, Wang D, Huang C, Zheng Y, Chen H, Pang CP, et al. Interactive effects of ATOH7 and RFTN1 in association with adult-onset primary open-angle glaucoma. Invest Ophthalmol Vis Sci. 2012;53(2):779-85. https://doi.org/10.1167/iovs.11-8277
13. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. Nature. 2008;454(7203):436-44. https://doi.org/10.1038/nature07205
14. Bayliss AL, Sundararaman A, Granet C, Mellor H. Raftlin is recruited by neuropilin-1 to the activated VEGFR2 complex to control proangiogenic signaling. Angiogenesis. 2020;23(3):371-83. https://doi.org/10.1007/s10456-020-09715-z
15. Deng H, Ji G, Ma J, Cai J, Cheng S, Cheng F. RNF19A inhibits bladder cancer progression by regulating ILK ubiquitination and inactivating the AKT/mTOR signalling pathway. Biol Direct. 2024;19(1):102. https://doi.org/10.1186/s13062-024-00562-2

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251637

Evaluation of ferroptosis markers and laboratory findings in patients with juvenile idiopathic arthritis

Seyda Dogantan[1]*, Burcu Bozkaya Yücel[2], Peren Perk[3], Arzu Şekerci Yüksel[4], Rahime Koç[1], Adem Keskin[5]

SUMMARY

OBJECTIVE: Juvenile idiopathic arthritis is the most prevalent chronic rheumatic disease of childhood. Ferroptosis is a distinct, regulated form of cell death that is iron-dependent and driven by the accumulation of lipid peroxides on cell membranes. The aim of this study was to measure the soluble transferrin receptor and malondialdehyde levels as ferroptosis markers in Turkish children with Juvenile idiopathic arthritis. We also evaluated the laboratory findings in the patients.

METHODS: Thirty children diagnosed with Juvenile idiopathic arthritis and 30 healthy children were included in the study. The sTfR1 and the malondialdehyde levels were measured by enzyme-linked immunosorbent assay.

RESULTS: The mean age of the patient group was 10.13±3.98, and the mean age of the control group was 10.67±3.21. Blood soluble transferrin receptor and malondialdehyde levels did not differ between the patient and control groups ($p>0.05$). In contrast, the p-value for the malondialdehyde results was statistically borderline. In the patient group, sedimentation rates were found to correlate with serum amyloid A, C-reactive protein levels. In addition, the highest malondialdehyde levels were found in systemic Juvenile idiopathic arthritis, and the highest soluble transferrin receptor levels were found in psoriatic arthritis subtypes. Both malondialdehyde and soluble transferrin receptor levels were lowest in those treated with methotrexate.

CONCLUSION: Although our findings show that soluble transferrin receptor and malondialdehyde levels did not differ between patient and control groups, larger-scale studies are needed to evaluate more specific biomarkers and/or targeted tissue sampling in different subtypes, controlling for confounding factors such as treatment. This could open new horizons for therapies targeting ferroptosis.

KEYWORDS: Juvenile idiopathic arthritis. Ferroptosis. Malondialdehyde. Methotrexate. ELISA.

INTRODUCTION

Juvenile idiopathic arthritis (JIA), which usually appears before the age of 16, is the most common rheumatic connective tissue disease of childhood. Characterized by leukocyte infiltration in the synovial fluid of the joints, this disease presents with chronic inflammation lasting at least 6 weeks and can lead to progressive damage to the joint tissue[1]. While dysregulation of the adaptive immune system primarily contributes to the pathophysiology in the oligoarticular and polyarticular subtypes of JIA, dysregulation of the innate immune system appears more prominent in the systemic form. This is particularly associated with increased levels of interleukin (IL)-1β, IL-6, and IL-18. Activation of monocytes and neutrophils, the key effector cells of joint inflammation and destruction, is one of the common pathogenic mechanisms in JIA[2]. In addition to the immune response, oxidative stress also plays a significant role in the pathogenesis of the disease. Oxidative stress results from an imbalance between the increase in reactive oxygen species (ROS) and reactive nitrogen species and the detoxification of these reactive intermediates or the repair of the protein, lipid, and DNA damage they cause[2].

In recent years, ferroptosis, a novel type of programmed cell death triggered by iron-dependent lipid peroxidation, has been identified. It is characterized by mitochondrial atrophy, increased membrane density, and changes in mitochondrial morphology[3]. An imbalance between oxidant and antioxidant systems, resulting from abnormal expression and activity of redox-active enzymes, can trigger ferroptotic cell death[4].

[1]Başakşehir Çam ve Sakura City Hospital, Department of Pediatric Rheumatology – Istanbul, Turkey.
[2]Samsun Research and Education Hospital, Department of Pediatric Rheumatology – Samsun, Turkey.
[3]Başakşehir Çam ve Sakura City Hospital, Department of Pediatric Neurology – Istanbul, Türkiye.
[4]Başakşehir Çam ve Sakura City Hospital, Department of Medical Biochemistry – İstanbul, Türkiye.
[5]Aydın Adnan Menderes University, Faculty of Medicine, Department of Medical Biochemistry – Aydin, Turkey.
*Corresponding author: drseydacayan@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on August 22, 2025. Accepted on February 23, 2026.
Scientific Editor: José Maria Soares Júnior

Ferroptosis; It is characterized by high lipid hydroperoxide levels and iron loading, and is a caspase and necrosome-independent cell death mechanism[5].

Increasing experimental evidence suggests that excessive iron accumulation can lead to oxidative tissue damage and organ dysfunction, contributing to the development of various diseases such as diabetes, cardiomyopathy, and cirrhosis. Karim and colleagues reported that excess iron disrupts cellular iron homeostasis, negatively impacts chondrocyte function, and leads to oxidative stress and cell death[6]. Furthermore, it has been shown that excess iron can exacerbate joint damage in osteoarthritic joints through increased ROS production and the accumulation of calcium phosphate, hydroxyapatite, and/or uric acid crystals[7].

Soluble transferrin receptor (sTfR) is the circulating form of transferrin receptor 1 (TfR1) found on the cell surface and plays a role in transporting iron into cells by binding to transferrin, the main iron-carrying protein in the blood[8]. On the other hand, malondialdehyde (MDA) is one of the most common and reliable biomarkers of lipid peroxidation and is frequently used in the assessment of oxidative stress.

In this study, we aimed to evaluate serum sTfR and MDA levels as markers associated with ferroptosis in Turkish children diagnosed with JIA. We also analyzed the patients' laboratory findings and examined the relationships between these parameters.

METHODS

Study design

The patient group consisted of 30 patients diagnosed with JIA followed up at the Pediatric Rheumatology Department of Başakşehir Çam and Sakura City Hospital. JIA diagnosis was made according to the International Union of Rheumatology Societies (ILAR) classification criteria[9]. Definitive diagnosis was confirmed in the presence of two primary criteria or one primary and two secondary criteria. None of the patients had additional chronic diseases. Clinical and laboratory findings of all patients were systematically evaluated. The control group consisted of 30 healthy children from the same geographic region, matched for age and gender, and without known chronic diseases.

In accordance with the ethical principles of the Helsinki Declaration, written informed consent was obtained from all participants and/or their legal guardians before inclusion in the study. Ethical approval for the study was granted by the Ethics Committee of Başakşehir Çam and Sakura City Hospital (Approval No: KAEK-11/10.09.2025.334).

Soluble transferrin receptor and malondialdehyde analysis

The sTfR and the MDA levels were measured by enzyme-linked immunosorbent assay (ELISA). These assessments were performed using the Sunred 201-12-7968 Human Soluble Transferrin Receptor ELISA kit and the Sunred 201-12-1952 Human MDA ELISA kit, respectively.

Statistical analysis

For statistical analyses, International Business Machines Statistical Package for the Social Sciences Statistics 22 (IBM Corp., Armonk, NY, USA) software was used for Windows. The normality of continuous variables was assessed using the Shapiro-Wilk test. It was determined that the data did not conform to a normal distribution. Therefore, continuous variables were expressed as median [25th percentile (Q1)–75th percentile (Q3)]. The Mann-Whitney U test was used for intergroup comparisons, and the chi-square test was used for comparisons of categorical variables. The Kruskal-Wallis test was used for comparisons of subgroups. Spearman correlation analysis was used to evaluate the relationship between variables. Bonferroni correction was applied for multiple comparisons. A statistical significance level of $p<0.05$ was accepted for all analyses.

RESULTS

Thirty patients aged 2–17 years, diagnosed with JIA, and 30 healthy children aged 6–17 years were included in the study. The mean age of the patient group was 10.13±3.98, and the mean age of the control group was 10.67±3.21. Of the patient group, 14 were women (46.67%) and 16 were men (53.33%). In the control group, these rates were 10 (33.33%) and 20 (67.67%), respectively. No significant difference was found between the two groups in terms of mean age and gender ratios ($p=0.570$ and $p=0.292$, respectively). Laboratory findings for the patient group are presented in Table 1. Blood sTfR and MDA levels did not differ between the patient and control groups. Erythrocyte sedimentation rate (ESR) and creatinine levels were found to be higher in the patient group than in the control group.

A correlation analysis was performed to evaluate the relationship between blood sTfR and MDA levels within each group. A significant positive correlation between blood sTfR and MDA levels was observed in both the patient ($r=0.700$) and control ($r=0.599$) groups ($p<0.001$). Additionally, a correlation test was performed to examine the relationship between the parameters found to be statistically significant between the two groups and other parameters. In the patient group, ESR

levels showed a positive correlation with serum amyloid A (r=0.570) and C-reactive protein (CRP) (r=0.725); these relationships remained significant even after Bonferroni correction (p=0.012 and p<0.001, respectively). In contrast, after Bonferroni correction, no significant correlation was found between creatinine levels and other parameters in the patient group. In the control group, no significant correlation was found between ESR and creatinine levels and other parameters after Bonferroni correction.

The median Juvenile Arthritis Disease Activity Score-27 (JADAS-27) score of the patient group was 4 (3–6), and the median disease duration was 9 (9–12) months. No significant correlation was found between JADAS-27 scores or disease duration and sTfR and MDA levels. The clinical characteristics of the patient group and the sTfR and MDA levels according to the clinical characteristics are presented in Table 2.

When evaluated in terms of JIA subtypes, the median MDA level was lowest in the Oligoarticular JIA subgroup and highest in the Systemic JIA subgroup. The median sTfR level was lowest in the Enthesitis-related arthritis subgroup and highest in the Psoriatic arthritis subgroup (Table 2).

In the treatment-based analysis, both MDA and sTfR levels were lowest in patients receiving Methotrexate treatment and highest in patients receiving Etanercept treatment. Regarding treatment duration, MDA levels were highest, while sTfR levels were lowest in patients receiving 12 months of treatment compared to other durations (Table 2).

However, there was no statistically significant difference in sTfR and MDA levels among subgroups created according to JIA subtypes, treatments applied, and treatment durations (Table 2).

Table 1. Laboratory findings of groups.

	Patient group (n=30)	Control group (n=30)	p
MDA	12.25 (9.45–17.44)	16.81 (12.8–22.87)	0.052
STfR	0.6 (0.55–0.73)	0.68 (0.51–0.97)	0.203
Serum amyloid A	0.56 (0.29–7.11)	0.65 (0.33–1.33)	0.749
Sedimentation	9 (4–20)	2.5 (1–4)	**<0.001**
CRP	0.5 (0.3–6.1)	1.2 (0.3–2.1)	0.976
Leucocyte	7.27 (6.06–8.56)	7.37 (6.1–9.19)	0.767
Hematocrit	36.4 (35–38.4)	38.5 (34.2–42.9)	0.109
Hemoglobin	12.1 (11.6–12.5)	12.6 (11–13.7)	0.437
Neutrophile	3.59 (2.58–5.4)	3.48 (2.78–4.06)	0.929
Lymphocyte	2.65 (2.21–3.38)	2.98 (2.49–3.6)	0.203
PLT	309 (245–369)	327.5 (287–363)	0.673
MCV	78 (74.7–80.1)	78.9 (77–82.8)	0.367
BUN	22.4 (18.2–28)	24 (22–26)	0.135
Creatinine	0.52 (0.42–0.64)	0.28 (0.28–0.28)	**<0.001**
ALT	15.5 (10–18)	14.5 (12–17)	0.923
AST	23 (19–28)	20 (17–28)	0.207
LDH	216 (193–257)	210 (182–240)	0.631
Vitamin B12	311.5 (242–389)	–	–
Vitamin D	22 (16–27)	–	–
Folic acid	6.72 (4.8–8.75)	–	–
Total protein	73 (70–76)	–	–
Albumin	46 (44–48)	–	–
T4	1.36 (1.24–1.54)	–	–
TSH	2.36 (1.49–2.62)	–	–
Calcium	9.69 (9.38–9.99)	–	–
Magnesium	2.07 (1.97–2.15)	–	–

MDA: malondialdehyde; sTfR: serum transferrin receptor 1; CRP: C-reactive protein; PLT: platelet; MCV: mean corpuscular volume; BUN: blood urea nitrogen ALT: alanine transaminase; AST: aspartate aminotransferase; LDH: lactate dehydrogenase; TSH: thyroid stimulating hormone. Statistically significant p-values are presented in bold.

DISCUSSION

The most common rheumatic disease in childhood, and one that leads to significant long-term morbidity, is JIA, a collection of various adolescent chronic arthritis syndromes. Despite various studies, the prevalence and incidence of JIA are still unknown due to the lack of standard classification techniques and regional differences in disease frequency[10]. The epidemiological characteristics of the disease are crucial to determining how hereditary and environmental variables affect disease progression. Furthermore, it can improve preventive health practices and provide clues to the best treatment method. Each region has a different JIA prevalence[11]. Abnormal activation of T cells, B cells, natural killer cells, dendritic cells, macrophages, and neutrophils, and the production of pro-inflammatory mediators leading to joint destruction and systemic problems, are part of the pathophysiological stage of JIA[12].

A recent meta-analysis examining 16 different studies reported that children with JIA have higher low-density lipoprotein cholesterol (LDL-C) levels compared to healthy children. Furthermore, it has been reported that dyslipidemia is more common in children with JIA than in healthy children, and that inflammatory conditions during the active phase of the disease or previous medication use may lead to lipid metabolism disorders in children with JIA[13]. Another study on lipid metabolism in children with JIA showed that dyslipidemia is common in JIA patients, particularly in the systemic subtype,

Table 2. The clinical characteristics of the patient group and the soluble transferrin receptor and malondialdehyde levels according to the clinical characteristics.

Clinical characteristics JIA patients (n=30)		MDA	p	STfR	p
JIA subtypes n (%)	Oligoarticular JIA 13 (43.33)	10.47 (9.05–16.18)	0.552	0.58 (0.53–0.65)	0.273
	Polyarticular JIA 6 (20)	14.98 (11.16–17.44)		0.67 (0.61–0.76)	
	Enthesitis-related arthritis 5 (16.67)	11.04 (10.95–12.68)		0.57 (0.56–0.58)	
	Psoriatic arthritis 4 (13.33)	19.38 (10.7–36.74)		0.83 (0.62–1.47)	
	Systemic JIA 2 (6.67)	19.81 (9.45–30.17)		0.69 (0.5–0.87)	
Treatment n (%)	Adalimumab 12 (40)	11.1 (9.61–17.32)	0.655	0.58 (0.53–0.65)	0.156
	Methotrexate 9 (30)	10.26 (9.05–21.13)		0.55 (0.5–0.67)	
	Etanercept 9 (30)	14.13 (11.81–16.18)		0.71 (0.63–0.76)	
Treatment duration n (%)	6 months 22 (73.33)	13.4 (10.47–17.06)	0.830	0.62 (0.56–0.73)	0.830
	9 months 4 (13.33)	10.04 (8.26–29.94)		0.62 (0.48–1.33)	
	12 months 4 (13.33)	15.29 (8.34–25.65)		0.57 (0.53–0.73)	

MDA: malondialdehyde; sTfR: serum transferrin receptor 1; JIA: Juvenile idiopathic arthritis.

and that this condition may be related to the underlying inflammatory process[14]. In contrast, MDA, a known biomarker of oxidative stress and end product of lipid peroxidation, is associated with chronic rheumatic diseases characterized by persistent inflammation, immune dysregulation, and progressive tissue damage; these processes are closely linked to increased ROS production. ROS contributes to synovial inflammation and to cartilage and bone destruction[15]. However, another recent study reported a positive correlation between MDA (a lipid peroxidation product) levels and LDL-C levels, and an association between MDA levels and dyslipidemia[16].

In this study, no significant difference was found in MDA levels between children with JIA and their healthy peers. This may be due to the fact that most of the children with JIA included in the study had the oligoarticular subtype. Indeed, in the evaluation between subtypes, the highest MDA levels were found in children with systemic JIA, while the lowest levels were found in children with the oligoarticular subtype.

Iron storage disorders and/or low hemoglobin levels are frequently observed in individuals with rheumatic diseases, including JIA. The high prevalence of anemia and iron deficiency in this patient group suggests that dysregulation in iron metabolism may play a significant role in the pathophysiology of the disease[17]. A recent study reported that hepcidin levels, an important iron-regulating hormone, correlated positively with the JADAS-27 score in children with JIA and may be useful in the early diagnosis and monitoring of disease exacerbations[18]. High hepcidin levels lead to iron retention in macrophages and other cells; consequently, serum iron decreases, and functional iron deficiency anemia may develop[19]. Another recent study suggested that clinicians use biomarkers such as sTfR and hepcidin for a more accurate assessment of body iron status. However, hepcidin measurement is not yet universally accessible, and sufficient standardization among methods has not been achieved[20].

In this study, no significant difference was found in sTfR levels between children with JIA and their healthy peers. Additionally, it was found that sTfR and MDA levels showed a positive correlation with each other, but no correlation with disease duration or JADAS-27 score. This may be related to the fact that the JIA children included in the study were under regular follow-up and did not have high disease activity. Indeed, the fact that the lowest sTfR levels were detected in JIA children who received the longest treatment also supports this view. Furthermore, metabolic reprogramming with methotrexate, commonly used in JIA patients, may mask or alter systemic ferroptosis biomarkers[21].

Both iron deficiency and iron excess are harmful to human health, because iron excess can trigger ferroptosis-mediated cell formation by increasing the proliferation of ROS[20]. Ferroptosis, iron-induced cell death, is characterized by iron accumulation along with excessive lipid peroxidation[22]. In this context, the effects of methotrexate treatment on metabolic and oxidative pathways are of great importance. Indeed, a recent metabolomic study by Tomioka et al. showed that methotrexate created a different metabolomic profile in JIA patients and significantly altered pathways associated with oxidative stress, lipid metabolism, and amino acid metabolism[21]. Similarly, a study investigating iron status during anti-tumor necrosis factor (TNF) treatment in children with JIA reported that sTfR was useful in assessing iron deficiency in children with JIA and that sTfR levels returned to normal with anti-TNF treatment[23].

In this study, both MDA and sTfR levels were found to be lowest in children with JIA who received methotrexate treatment. Additionally, other patients were receiving anti-TNF therapy. One reason for the lack of difference in MDA and sTfR levels between children with JIA and their healthy peers may be related to the treatment.

This study has some limitations. First, the relatively small sample size may have limited statistical power, particularly in subgroup analyses, and reduce the generalizability of the results. This may have increased the risk of Type II error due to the failure to detect some true differences. Second, the heterogeneous nature of the treatments administered in the patient group (treatment confounding effect) may have created a potential confounding effect at biomarker levels. Furthermore, it should be considered that the biomarkers evaluated have limited disease specificity and may be affected by different pathophysiological processes such as inflammation or oxidative stress. Therefore, the findings should be interpreted cautiously and supported by multicenter studies with larger, more homogeneous treatment groups. However, the preliminary data obtained in our study can be considered a strength, as they provide a reference and hypothesis-forming basis for future research with larger and more homogeneous sample sizes.

CONCLUSION

In conclusion, although no systemic increase in sTfR and MDA levels was detected in this cohort followed under treatment, these findings do not rule out a possible role of ferroptosis in the pathophysiology of JIA. The absence of significant changes may be related to controlled disease activity and ongoing treatments. Therefore, the contribution of ferroptosis to JIA, especially in the context of specific disease stages or subtypes, remains an open question. Better-designed studies controlling for significant confounding factors such as methotrexate treatment are needed. Furthermore, given that sTfR and MDA are indirect markers, studies including glutathione peroxidase 4 (GPX4) activity measurement, comprehensive lipidomic analyses, and/or targeted tissue sampling (e.g., synovial fluid) are needed to allow for a more direct assessment of ferroptosis.

AUTHORS' CONTRIBUTIONS

SD: Conceptualization, Methodology, Resources, Software, Writing – original draft, Writing – review & editing. **BBY:** Conceptualization, Validation. **PP:** Data curation, Formal analysis, Investigation, Resources. **AŞY:** Investigation, Software. **RK:** Data curation, Formal analysis, Resources. **AK:** Formal analysis, Writing – original draft, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Zurawa-Janicka D, Renke J, Popadiuk S, Skorko-Glonek J, Szumera M, Plata-Nazar K, et al. Preferential immunoglobulin oxidation in children with juvenile idiopathic arthritis. Scand J Rheumatol. 2006;35(3):193-200. https://doi.org/10.1080/03009740500492976
2. Bella S, Rinaldi M, Ludovico A, Donato G, Donato G, Salpietro V, et al. Genetic background and molecular mechanisms of Juvenile idiopathic arthritis. Int J Mol Sci. 2023;24(3):1846. https://doi.org/10.3390/ijms24031846
3. Cao JY, Dixon SJ. Mechanisms of ferroptosis. Cell Mol Life Sci. 2016;73(11-12):2195-209. https://doi.org/10.1007/s00018-016-2194-1
4. Tang D, Chen X, Kang R, Kroemer G. Ferroptosis: molecular mechanisms and health implications. Cell Res. 2021;31(2):107-25. https://doi.org/10.1038/s41422-020-00441-1
5. Friedmann Angeli JP, Schneider M, Proneth B, Tyurina YY, Tyurin VA, Hammond VJ, et al. Inactivation of the ferroptosis regulator Gpx4 triggers acute renal failure in mice. Nat Cell Biol. 2014;16(12):1180-91. https://doi.org/10.1038/ncb3064
6. Karim A, Bajbouj K, Shafarin J, Qaisar R, Hall AC, Hamad M. Iron overload induces oxidative stress, cell cycle arrest and apoptosis in chondrocytes. Front Cell Dev Biol. 2022;10:821014. https://doi.org/10.3389/fcell.2022.821014
7. Naughton DP. Iron(III)-mediated intra-articular crystal deposition in arthritis: a therapeutic role for iron chelators. Med Hypotheses. 2001;57(1):120-2. https://doi.org/10.1054/mehy.2000.1174
8. Sawicki KT, Jesus A, Ardehali H. Iron metabolism in cardiovascular disease: physiology, mechanisms, and therapeutic targets. Circ Res. 2023;132(3):379-96. https://doi.org/10.1161/CIRCRESAHA.122.321667
9. Petty RE, Southwood TR, Manners P, Baum J, Glass DN, Goldenberg J, et al. International league of associations for rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. J Rheumatol. 2004;31(2):390-2. PMID: 14760812
10. Giancane G, Consolaro A, Lanni S, Davì S, Schiappapietra B, Ravelli A. Juvenile idiopathic arthritis: diagnosis and treatment. Rheumatol Ther. 2016;3(2):187-207. https://doi.org/10.1007/s40744-016-0040-4
11. Huang HYR, Wireko AA, Miteu GD, Khan A, Roy S, Ferreira T, et al. Advancements and progress in juvenile idiopathic arthritis: a review of pathophysiology and treatment. Medicine

(Baltimore). 2024;103(13):e37567. https://doi.org/10.1097/MD.0000000000037567

12. Zaripova LN, Midgley A, Christmas SE, Beresford MW, Baildam EM, Oldershaw RA. Juvenile idiopathic arthritis: from aetiopathogenesis to therapeutic approaches. Pediatr Rheumatol Online J. 2021;19(1):135. https://doi.org/10.1186/s12969-021-00629-8
13. Zhao WJ, Deng JH, Li CF. Lipid profiles in patients with juvenile idiopathic arthritis: a systematic literature review and meta-analysis. Lipids Health Dis. 2023;22(1):136. https://doi.org/10.1186/s12944-023-01885-1
14. Rodrigues WDR, Sarni ROS, Fonseca FLA, Araújo A, Len CA, Terreri MT. Biomarkers of lipid metabolism in patients with juvenile idiopathic arthritis: relationship with disease subtype and inflammatory activity. Pediatr Rheumatol Online J. 2021;19(1):66. https://doi.org/10.1186/s12969-021-00538-w
15. Cordiano R, Gioacchino M, Mangifesta R, Panzera C, Gangemi S, Minciullo PL. Malondialdehyde as a potential oxidative stress marker for allergy-oriented diseases: an update. Molecules. 2023;28(16):5979. https://doi.org/10.3390/molecules28165979
16. Chittaragi GH, Khubya D, Ganesh V. Linking serum malondialdehyde with dyslipidemia in patients with metabolic syndrome. Bioinformation. 2025;21(6):1693-8. https://doi.org/10.6026/973206300211693
17. Pecher AC, Bach S, Pauluschke-Fröhlich J, Abele H, Henes J, Henes M. Anemia and iron deficiency in pregnant women with rheumatic diseases. Joint Bone Spine. 2024;91(1):105650. https://doi.org/10.1016/j.jbspin.2023.105650
18. Troiani R, Bella S, Muselli M, D'Egidio M, Sassano G, Ludovico A, et al. Serum hepcidin evaluation as a promising biomarker in juvenile idiopathic arthritis. Clin Exp Rheumatol. 2025;43(1):151-9. https://doi.org/10.55563/clinexprheumatol/a51j2e
19. Nemeth E, Ganz T. Hepcidin-ferroportin interaction controls systemic iron homeostasis. Int J Mol Sci. 2021;22(12):6493. https://doi.org/10.3390/ijms22126493
20. Girelli D, Marchi G, Busti F, Chesini F, Castagna A. Diagnostics: markers of body iron status. Adv Exp Med Biol. 2025;1480:387-98. https://doi.org/10.1007/978-3-031-92033-2_25
21. Tomioka RB, Ferreira GRV, Aikawa NE, Maciel GAR, Junior JMS, Baracat EC, et al. Metabolomics in juvenile idiopathic arthritis: a distinct profile in patients under methotrexate. Clinics (Sao Paulo). 2025;80:100522. https://doi.org/10.1016/j.clinsp.2024.100522
22. Li J, Jia YC, Ding YX, Bai J, Cao F, Li F. The crosstalk between ferroptosis and mitochondrial dynamic regulatory networks. Int J Biol Sci. 2023;19(9):2756-71. https://doi.org/10.7150/ijbs.83348
23. Aalto K, Honkanen V, Lahdenne P. Iron status during anti-TNF therapy in children with juvenile idiopathic arthritis. Clin Rheumatol. 2011;30(1):115-9. https://doi.org/10.1007/s10067-010-1571-7

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251739

Examining the relationship between breastfeeding duration at age two and beyond with parenting styles and infant feeding attitudes in Turkey (2023–2024)

Şenay Türe[1]*, İlknur Münevver Gönenç[2], Sevtap Velipaşaoğlu[1]

SUMMARY

OBJECTIVE: Exclusive breastfeeding for 6 months and continued breastfeeding for 24 months remain low due to various challenges. The aim of this study was to examine the impact of mothers' infant feeding attitudes and parenting styles, as modifiable factors, on breastfeeding duration.

METHODS: This cross-sectional descriptive study was conducted at Akdeniz University Faculty of Medicine Hospital between April 1, 2023, and April 1, 2024. The sample comprised 224 mothers of children aged 2–3 years who attended the child health follow-up clinic. Data were obtained through face-to-face interviews using the Introductory Information Form, Child Nutrition Questionnaire, Iowa Infant Feeding Attitudes Scale, and Parenting Styles Scale. Group differences were examined with t-tests and analysis of variance, correlations with Pearson's coefficient, and categorical data with chi-square tests. A binary logistic regression evaluated the effects of parenting style and feeding attitudes on breastfeeding beyond 2 years.

RESULTS: While 98.7% of participants breastfed at least once, only 38.8% exclusively breastfed for the first 6 months, and 40.6% continued breastfeeding for 24 months or longer. It was found that a one-unit increase in the Iowa Infant Feeding Attitudes Scale score increased the probability of breastfeeding beyond 24 months by 5.4%, while each unit increase in the authoritarian attitude decreased it by 9.7%. Formula feeding duration was also positively correlated with authoritarian and overprotective parenting styles.

CONCLUSION: Parenting styles and mothers' infant feeding attitudes are two important factors affecting the rates of exclusive breastfeeding for the first 6 months and continuing breastfeeding beyond 24 months.

KEYWORDS: Breastfeeding. Exclusive. Parenting. Infant nutrition. Time factors.

INTRODUCTION

In 2012, the World Health Organization (WHO) set a goal for 50% of infants under 6 months to be exclusively breastfed by 2025[1]. Yet, most countries remain below these benchmarks. Marked disparities exist across high-income nations: exclusive breastfeeding rates at 6 months are ~1% in the United Kingdom, 13% in Denmark, 39% in the Netherlands, and 41% in Turkey[2].

Multiple factors hinder the achievement of breastfeeding goals. Sociodemographic variables, such as maternal age, education, and socioeconomic status, strongly influence infant feeding, yet these factors are often difficult to alter[3,4]. Therefore, interventions should emphasize modifiable determinants. Targeted interventions and support services can substantially promote success, but first require identifying women who may choose not to breastfeed.

Maternal feeding attitudes have been extensively examined, with the Iowa Infant Feeding Attitudes Scale (IIFAS) validated across diverse contexts. However, prior studies primarily address feeding within the first 6–12 months[4-6]. Despite WHO recommendations to continue breastfeeding for 2 years or longer, to our knowledge, IIFAS has not yet been applied to extended breastfeeding durations.

Previous research has demonstrated an association between parenting styles and breastfeeding behaviors[7-9]. These parenting styles are categorized as democratic, authoritarian, permissive, and overprotective. It is essential to find answers to questions such as how parenting styles, especially aspects related to control and care, may affect or guide the decision to breastfeed or formula feed[8,10-12]. Studies in this area are few and typically cover the first 6 months or 1 year[8,9]. Despite WHO recommendations to continue breastfeeding

[1]Akdeniz University, School of Medicine, Department of Pediatrics, Division of Social Pediatrics – Antalya, Turkey.
[2]Ankara University, Faculty of Nursing – Ankara, Turkey.
*Corresponding author: senayture@akdeniz.edu.tr, drsenayture@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on October 10, 2025. Accepted on December 23, 2025.
Scientific Editor: José Maria Soares Júnior

for two years or longer, to our knowledge, we did not find any studies covering this period. Therefore, there is a need for research examining which parenting styles affect mothers' feeding attitudes and behaviors, accounting for breastfeeding durations of 2 years or longer.

Despite WHO recommendations, breastfeeding rates for children 2 years of age and beyond remain very low[2]. Therefore, it is a priority for researchers and policymakers to identify the risk factors by determining the reasons for cessation or non-initiation of breastfeeding and to provide support to women who want to breastfeed their children. We wanted to investigate the question: Do parenting styles and infant feeding attitudes affect breastfeeding duration at 2 years of age and beyond?

Purpose

In this study, we aimed to examine the relationships among breastfeeding duration through 2 years of age and beyond, and infant feeding attitudes and parenting styles, which are more flexible, "modifiable" factors for interventions.

METHODS

This cross-sectional, correlational study was conducted at the Pediatrics and Social Pediatrics Outpatient Clinics of Akdeniz University Hospital between April 1, 2023, and April 1, 2024. Participants were mothers of healthy children aged 2–3 years attending routine follow-up. Eligibility required literacy in Turkish and effective communication, while exclusions applied to mothers or infants with breastfeeding-related medical conditions, psychiatric diagnoses, or prematurity.

Calculating the sample size

The G*Power 3.1.9.4 program was used to calculate the sample size. In the sample article, the mean Parenting Styles Scale (PSS) scores for the two groups were 37.48±5.45 and 34.10±5.55, respectively, and the effect size was 0.61[13]. Using the data from this study, with an α error value of 0.05, a power (1-β error value) of 0.95, an effect size of 0.61, and an N1/2 value of 1, a total of 118 cases were needed for children aged two and over, with at least 59 individuals in the breastfed and non-breastfed groups. A total of 224 individuals were reached in the study: 91 breastfed children aged two and over, and 113 non-breastfed children.

Measurement

Data were collected using four instruments: the Introductory Information Form, the Child Nutrition Questionnaire, the IIFAS, and the PSS.

The Introductory Information Form (18 items) assessed maternal sociodemographic and basic child characteristics. The Child Nutrition Questionnaire (nine items) evaluated children's nutritional status from birth.

IIFAS; the scale, developed by De La Mora and Russell and adapted to Turkish by Eksioğlu et al., measures women's attitudes toward breastfeeding and feeding method choice[5,14]. It consists of 17 items on a five-point Likert scale, with reverse scoring for formula-feeding items. Scores range from 17 to 85, with higher scores indicating a positive attitude toward breastfeeding. The Turkish version had a Cronbach Alpha of 0.71 in Eksioğlu et al.'s study[14]. In this study, Cronbach's α value was found to be 0.703.

PSS; the scale, developed in Turkish by Karabulut Demir and Şendil, assesses parenting attitudes of mothers and fathers with children aged 2–6. It consists of 46 items across four sub-dimensions, with scores ranging from 1 (never like this) to 5 (always like this)[10]. Higher scores indicate stronger adoption of the attitude in that dimension. Cronbach's alpha values in the study were 0.83 for the "democratic" dimension, 0.76 for "authoritarian," 0.75 for "overprotective," and 0.74 for "permissive." In this study, the following values were found to be 0.86, 0.82, 0.81, and 0.71, respectively. Mothers participating in the study were not categorized according to a single characteristic, but received one point for each characteristic identified by the PSS[10].

Data collection

Participants were informed about the study, and written consent was obtained prior to their participation. Data were collected face-to-face by the same physician in a private outpatient room to ensure comfort and confidentiality. Completion of the forms required approximately 15–20 min.

Data analysis

Statistical analyses were conducted using Statistical Package for the Social Sciences 20.0. The Shapiro-Wilk test was employed to assess normality, and variables with skewness within ±1.5 were subjected to parametric tests. Group differences were evaluated using Independent Samples t-tests and analysis of variance, while correlations were determined through Pearson's correlation analysis. Categorical variables were analyzed using chi-square tests, with Pearson or Fisher's Exact tests applied as appropriate. Statistical significance was set at $p<0.05$. Binary logistic regression was utilized to examine the effects of parenting style and feeding attitudes on breastfeeding duration beyond 2 years.

RESULTS

Of the 224 mothers who participated in the study, 70.5% were university graduates, and 41.5% were working. Among participants, 98.7% breastfed at least once. Of the children, 38.8% were exclusively breastfed for the first 6 months, and 23.7% were given water. After 6 months, 70.5% were introduced to complementary foods. Breastfeeding continued as follows: 83.5% between 6 and 12 months, 71% between 12 and 18 months, 54.9% between 18 and 24 months, and 40.6% for 24 months or more. Those exclusively breastfed for 6 months were 80% more likely to continue breastfeeding after 24 months (odds ratio: 1.80, 95%CI 1.313–2.466, p<0.0001).

The distribution of participants' IIFAS and PSS mean scores according to breastfeeding characteristics is given in Table 1. The IIFAS mean score was higher in those breastfed for 6 months or more compared to those not breastfed (p<0.05). Additionally, those who received breast milk with supplementary food between 6 and 12 months had higher scores than those given formula (p<0.001). Those who exclusively breastfed for six months had lower authoritarian parenting scores (p=0.021). Those who gave water in the first 6 months had higher overprotective parenting scores (p=0.039). Mothers who breastfed for 24 months or more had lower authoritarian scores (p=0.011). Significant differences in overprotective parenting scores were found between feeding patterns between 6 and 12 months (p=0.033), with higher scores for those combining breast milk, formula, and complementary feeding. A significant difference in authoritarian scores was found after 24 months (p=0.013).

A low-level, significantly positive correlation was found between the participants' IIFAS score and breastfeeding duration, as well as between the democratic parenting attitude score. Additionally, there was a low-level negative correlation between

Table 1. Distribution of participants' parenting style scale sub-dimension scores and Iowa Infant Feeding Attitudes Scale score averages by breastfeeding characteristics.

Categories	N/%	IIFAS Mean (SD)	p	Permissive Mean (SD)	p	Democratic Mean (SD)	p	Authoritarian Mean (SD)	p	Overprotective Mean (SD)	p
Exclusive breastfeeding for the first 6 months											
Yes	87/38.8	67.51 (6.00)	**<0.001**	22.12 (4.60)	0.969	75.20 (6.47)	0.177	19.26 (4.67)	**0.021**	31.62 (6.18)	0.248
No	137/61.2	64.47 (6.63)		22.14 (4.41)		73.85 (7.61)		20.93 (5.98)		32.61 (6.29)	
Provided water to the baby during the first 6 months											
Yes	53/23.7	63.81 (6.49)	**0.019**	21.68 (4.50)	0.401	74.43 (8.26)	0.944	20.68 (5.78)	0.551	33.77 (6.89)	**0.039**
No	171/76.3	66.22 (6.48)		22.27 (4.47)		74.35 (6.87)		20.15 (5.50)		31.74 (5.98)	
Breastfeeding status over 24 months											
Yes	91/40.6	67.04 (5.69)	**0.006**	21.92 (4.37)	0.568	74.84 (7.38)	0.433	19.12 (4.86)	**0.011**	31.31 (6.56)	0.075
No	133/59.4	64.70 (6.93)		22.27 (4.55)		74.06 (7.10)		21.06 (5.87)		32.84 (5.98)	
Feeding pattern between 6 and 12 months											
B+CF (a)	152/67.9	66.86 (6.05)	**<0.001** difference (a–c)	21.93 (4.44)	0.478	74.48 (7.08)	0.767	19.84 (4.95)	0.256	31.70 (6.17)	**0.033** difference (a–b)
B+F+CF (b)	35/15.6	64.51 (6.29)		22.29 (4.13)		75.06 (7.31)		21.79 (6.87)		34.56 (6.37)	
F+CF (c)	32/14.3	61.91 (7.54)		23.13 (5.01)		73.41 (7.72)		20.91 (6.20)		27.80 (4.02)	
CF (d)	5/2.2	61.00 (4.89)		20.60 (4.21)		72.80 (8.43)		19.20 (8.25)		32.23 (6.25)	
Feeding pattern over 24 months											
B+CF (a)	90/40.4	67.12 (5.67)	**0.006** difference (a–c)	21.91 (4.40)	0.694	74.72 (7.35)	0.447	19.19 (4.84)	**0.013**	31.16 (6.44)	0.137
F+CF (b)	22/9.8	62.64 (7.74)		22.82 (4.74)		72.55 (9.16)		22.82 (6.89)		33.23 (6.45)	
CF (c)	111/49.8	65.11 (6.72)		22.16 (44.53)		74.36 (6.62)		20.71 (5.61)		32.76 (5.91)	

N: number of people; %: frequency; SD: standard deviation; p: significance level; IIFAS: Iowa Infant Feeding Attitude Scale; B: breastfeeding; F: formula; CF: complementary food. Statistically significant p-values are indicated in bold.

formula feeding duration and the overprotective parenting attitude score. A low-level positive relationship was also observed between formula feeding duration and both authoritarian and overprotective parenting attitude scores (Table 2).

Binary logistic regression analysis was conducted to assess the impact of parenting styles and infant feeding behaviors on breastfeeding status beyond 24 months. Maternal age, education level, family income, maternal employment status, and the presence of a sibling were included in the model and controlled for as potential confounding variables (Table 3). It was found that a one-unit increase in the IIFAS score increased the probability of breastfeeding beyond 24 months by 5.4%, while each unit increase in the authoritarian attitude decreased it by 9.7%.

DISCUSSION

Breast milk is the optimal infant nutrition, and its promotion remains a global public health priority[1]. This study, conducted in a baby-friendly hospital, found that while 98.7% of mothers initiated breastfeeding, continuation declined to 83.5% at 6 months, 71% at 12 months, and 40% at 24 months. Exclusive breastfeeding for 6 months reached 38.8%, exceeding rates in some European countries yet remaining below WHO targets[1,2]. These figures should also be interpreted cautiously, as the sample included only healthy mothers and infants.

Consistent with prior research, maternal attitudes were strongly associated with breastfeeding during the first 6 months[3,6]. Beyond this period, we found that mothers who breastfed for 6–24 months or longer had significantly higher IIFAS scores than mothers who did not breastfeed at all. We also showed that each unit increase in IIFAS value increased the probability of breastfeeding at age 2 and beyond by 5.4%. In our study, higher IIFAS scores were associated with longer breastfeeding and reduced formula use. These results align with prior evidence linking positive maternal attitudes to breastfeeding initiation and duration, while also contributing novel data on practices extending past 2 years[6,15].

Table 2. Correlation table some characteristics of the participants.

Variable	1	2	3	4	5	6	7	8
1. IIFAS								
r (p-value)	1							
2. Breastfeeding duration/day*								
r (p-value)	0.283 (<0.001)	1						
3. Feeding duration/day								
r (p-value)	-0.233 (<0.001)	-0.607 (<0.01)	1					
4. Timing of complementary feeding**								
r (p-value)	0.122 (0.068)	0.151 (0.024)	-0.081 (0.228)	1				
5. Permissive								
r (p-value)	-0.068 (0.316)	-0.029 (0.664)	0.093 (0.169)	0.046 (0.497)	1			
6. Democratic								
r (p-value)	0.221 (<0.001)	0.072 (0.291)	-0.035 (0.604)	-0.019 (0.784)	-0.077 (0.256)	1		
7. Authoritarian								
r (p-value)	-0.023 (0.732)	-0.113 (0.094)	0.139 (0.039)	-0.045 (0.505)	0.191 (0.004)	-0.430 (<0.001)	1	
8. Overprotective								
r (p-value)	-0.163 (0.015)	-0.087 (0.198)	0.150 (0.026)	-0.082 (0.227)	0.172 (0.010)	0.171 (0.010)	0.016 (0.817)	1

r: Pearson correlation coefficient; p: Level of significance; IIFAS: Iowa Infant Feeding Attitude Scale. *Breastfeeding duration/day=Expresses the baby's breastfeeding duration in days (e.g., 90 days, 350 days, etc.). **Timing of complementary feeding=It indicates the day on which complementary feeding was initiated (e.g., complementary feeding started on the 120th or 181st day of the baby).

Table 3. Evaluation of the effects of parenting styles and Iowa Infant Feeding Attitudes Scale score on breastfeeding rates in children aged 2 and over using binary logistic regression analysis.

Independent variables		N/%	Adjusted odds ratio (95%CI)	p
Mother's age			1.096 (1.02–1.17)	**0.008**
Mother's education level	High school or less	66/29.5	1*	0.893
	University	158/70.5	1.057 (0.47–2.37)	
Mother's working status	Not working	95/42.6	1*	
	Working	128/57.4	0.955 (0.48–1.89)	0.896
Perceived income level	Income is less than expenses	44/19.6	1*	0.669
	Balanced income and expenses	139/62.1	0.825 (0.37–1.79)	
	Income is more than expenses	41/18.3	1.160 (0.41–3.23)	
Sibling presence	No	106/47.3	1*	0.622
	Yes	118/52.7	1.174 (0.62–2.22)	
IIFAS score			1.054 (1.00–1.10)	**0.031**
Permissive score			1.021 (0.95–1.09)	0.570
Democratic score			0.982 (0.93–1.02)	0.443
Authoritarian score			0.903 (0.84–0.96)	**0.002**
Overprotective score			0.977 (0.92–1.03)	0.405

Nagelkerke R Square: 0.151. *Reference group. p: level of significance; IIFAS: Iowa Infant Feeding Attitude Scale; CI: confidence interval. Statistically significant p-values are indicated in bold.

Parenting styles have also been associated with breastfeeding attitudes[8,9,16]. In their study involving infant-mother pairs in the first year of life, Brown and Arnott showed that a high level of parent-led routine reduced the likelihood of a mother initiating and continuing breastfeeding[8]. In the study conducted by Davis et al., it was shown that higher scores in the "permissive" and "authoritative/competent" dimensions were significantly associated with more positive attitudes toward breastfeeding, and higher scores in the "uninvolved" and "authoritarian" dimensions were associated with less positive attitudes[9].

This study extends the literature from 1 to 2 years to examine the impact of parenting styles on breastfeeding duration beyond 2 years. We found that each unit increase in authoritarian parenting style decreased the likelihood of breastfeeding beyond 2 years by 9.7%, while each unit increase in IIFAS score increased it by 5.4%. There was also a positive association between breastfeeding duration and IIFAS scores and a democratic parenting style, and a negative association with an overprotective parenting style.

In this study, mothers who did not exclusively breastfeed their babies for the first 6 months had higher authoritarian parenting style scores than mothers who did. We found that mothers who gave their babies water and formula in addition to breastmilk had higher overprotective parenting style scores. The relationship between parenting styles and breastfeeding attitudes found in this study is consistent with studies in this field[8,9,16].

It is not easy to establish a direct relationship between parenting styles and breastfeeding and breastfeeding duration. Breastfeeding for a long time may affect the bond between mother and baby and may change the parenting style, and it has been noted that mothers with a certain parenting style may show variability in breastfeeding and breastfeeding duration[8,11,12,16].

Our study results are consistent with the literature showing that modifiable factors such as mothers' infant feeding attitudes and parenting styles are associated with breastfeeding duration and contribute to the knowledge in this field by providing data for 6 months later[3,6,8,9].

These findings may guide health professionals in prenatal and postnatal care by linking parenting styles with breastfeeding attitudes, enabling early identification of mothers requiring additional counseling. Recognizing mothers with low infant feeding attitude scores and authoritarian or overprotective parenting styles allows health providers, such as midwives, nurses, and doctors, to offer personalized guidance and solutions to meet each mother's specific needs.

Strengths and weaknesses of the study

This study has several limitations. First, it included only mothers of children aged 2–3 years, and retrospective questions about birth and infant feeding may have introduced recall and social desirability biases. However, there is evidence that maternal recall of breastfeeding practice up to 3 years is valid

and reliable[17]. Second, the cross-sectional design limits causal inference and does not capture long-term changes in parenting attitudes. Despite these limitations, the study provides valuable insight into the relationship between parenting attitudes and breastfeeding through 2 years, offering a basis for future longitudinal research.

In future studies, health practitioners can design interventions for mothers who may feel negatively about breastfeeding by understanding the determinants of positive attitudes toward breastfeeding. A practical approach is to clarify mothers' questions about breastfeeding by assessing infant feeding attitudes, parenting style, and expected breastfeeding duration. These interventions may facilitate breastfeeding initiation and increase breastfeeding rates through age 2, thereby improving overall maternal and infant health.

CONCLUSION

Parenting styles and infant feeding attitudes are two important factors affecting exclusive breastfeeding rates for the first 6 months and continuing breastfeeding beyond 24 months, which are the WHO targets. Identifying mothers with low infant feeding attitude scale scores and authoritarian or overprotective parenting styles can contribute to the individualization of information provided by health professionals, such as midwives, nurses, and doctors, who offer counseling and the ability to provide solutions to the mother's problems.

AUTHORS' CONTRIBUTIONS

ŞT: Conceptualization, Data curation, Formal Analysis, Writing – original draft. **İMG:** Conceptualization, Formal Analysis, Writing – original draft. **SV:** Conceptualization, Writing – original draft.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. World Health Organization. Breastfeeding recommendations. 2012. [cited on 2024 Nov 27]. Available from: https://www.who.int/health-topics/breastfeeding#tab=tab_2
2. Theurich MA, Davanzo R, Busck-Rasmussen M, Díaz-Gómez NM, Brennan C, Kylberg E, et al. Breastfeeding rates and programs in Europe: a survey of 11 national breastfeeding committees and representatives. J Pediatr Gastroenterol Nutr. 2019;68(3):400-7. https://doi.org/10.1097/MPG.0000000000002234
3. Scott JA, Binns CW, Oddy WH, Graham KI. Predictors of breastfeeding duration: evidence from a cohort study. Pediatrics. 2006;117(4):e646-55. https://doi.org/10.1542/peds.2005-1991
4. Thulier D, Mercer J. Variables associated with breastfeeding duration. J Obstet Gynecol Neonatal Nurs. 2009;38(3):259-68. https://doi.org/10.1111/j.1552-6909.2009.01021.x
5. Mora A, Russell DW. The Iowa Infant Feeding Attitude Scale: analysis of reliability and validity. J Appl Soc Psychol. 1999;29:2362-80.
6. Cox KN, Giglia RC, Binns CW. The influence of infant feeding attitudes on breastfeeding duration: evidence from a cohort study in rural Western Australia. Int Breastfeed J. 2015;10:25. https://doi.org/10.1186/s13006-015-0048-3
7. Blissett J, Haycraft E. Are parenting style and controlling feeding practices related? Appetite. 2008;50(2-3):477-85. https://doi.org/10.1016/j.appet.2007.10.003
8. Brown A, Arnott B. Breastfeeding duration and early parenting behaviour: the importance of an infant-led, responsive style. PLoS One. 2014;9(2):e83893. https://doi.org/10.1371/journal.pone.0083893
9. Davis AMB, Coleman C, Kramer RSS. Parenting styles and types: breastfeeding attitudes in a large sample of mothers. Midwifery. 2021;103:103142. https://doi.org/10.1016/j.midw.2021.103142
10. Demir Karabulut E, Şendil G. Parents' Attitude Scale (PAS). Türk Psikoloji Yazıları. 2008;11:15-25.
11. Silva GA, Costa KA, Giugliani ER. Infant feeding: beyond the nutritional aspects. J Pediatr (Rio J). 2016;92(3 Suppl 1):S2-7. https://doi.org/10.1016/j.jped.2016.02.006
12. Weaver JM, Schofield TJ, Papp LM. Breastfeeding duration predicts greater maternal sensitivity over the next decade. Dev Psychol. 2018;54(2):220-7. https://doi.org/10.1037/dev0000425
13. Aydoğdu F, Dilekmen M. The evaluation of parents' attitudes in terms of different variables. Bayburt Fac Educ J. 2016;11:2
14. Eksioğlu A, Yeşil Y, Çeber Turfan E. The translation and validation of the Iowa Infant Feeding Attitude Scale into Turkish. J Educ Res Nurs. 2016;13:209-15.
15. Bai Y, Middlestadt SE, Peng CY, Fly AD. Predictors of continuation of exclusive breastfeeding for the first six months of life. J Hum Lact. 2010;26(1):26-34. https://doi.org/10.1177/0890334409350168
16. Hubbs-Tait L, Kennedy TS, Page MC, Topham GL, Harrist AW. Parental feeding practices predict authoritative, authoritarian, and permissive parenting styles. J Am Diet Assoc. 2008;108(7):1154-61;discussion 1161-2. https://doi.org/10.1016/j.jada.2008.04.008
17. Li R, Scanlon KS, Serdula MK. The validity and reliability of maternal recall of breastfeeding practice. Nutr Rev. 2005;63(4):103-10. https://doi.org/10.1111/j.1753-4887.2005.tb00128.x

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251769

Assessing the accuracy and consistency of responses from Chat Generative Pre-trained Transformer in human papillomavirus infection and protection

Oguz Arslan[1]*, Burak Giray[2]

SUMMARY

OBJECTIVE: Chat Generative Pre-trained Transformer is an artificial intelligence chat system that is pre-trained on a large dataset of language data, providing detailed answers to user inquiries. The aim of this study was to evaluate the accuracy and consistency of the model's answers to frequently asked patient questions about human papillomavirus infection.

METHODS: The questions prepared for input into Chat Generative Pre-trained Transformer were sourced from globally recognized associations and organizations and were divided into four categories: basic knowledge, diagnosis, treatment, and preventive medicine. The evaluation levels consisted of comprehensive, correct but inadequate, some correct and some incorrect, and completely incorrect. Three independent reviewers evaluated the responses.

RESULTS: A total of 123 questions about human papillomavirus infection were prepared. Of the questions prepared for entry into Chat Generative Pre-trained Transformer related to human papillomavirus infection, 52% were categorized as preventive medicine, 28.5% as basic knowledge, 13.8% as diagnosis, and 5.7% as treatment. The model provided "comprehensive" answers to 115 out of 123 (93.5%) questions. When examined by category, the model provided comprehensive answers to 94.3% of the questions related to "basic knowledge," 88.2% of the questions about "diagnosis," 100% of the questions about "treatment," and 93.8% of the questions about "preventive medicine." Additionally, the model provided reproducible answers to 120 (97.6%) questions.

CONCLUSION: Chat Generative Pre-trained Transformer performed well by providing accurate and consistent answers to frequently asked questions about human papillomavirus infection risk factors, transmission, diagnosis, impacts, associated cancers, prevention methods, and vaccination. The model may be considered a valuable informational resource for patients when used in conjunction with care provided by healthcare professionals.

KEYWORDS: Artificial intelligence. ChatGPT. Human papillomavirus. Pap smear. Vaccination.

INTRODUCTION

Human papillomavirus (HPV) is a common infection primarily transmitted through sexual contact and causes cellular changes in the oral mucosa and genital areas[1]. Nearly 80% of sexually active women get an HPV infection at some point in their lives. The natural immune system responds to and clears approximately 90% of HPV infections, preventing them from becoming persistent[2]. Cervical cancer, arising from HPV infection, is the second most prevalent cancer among women under 50 and the fourth most prevalent across all age groups[1].

Chat Generative Pre-trained Transformer (ChatGPT) is an artificial intelligence (AI) chatbot that became available to people on November 30, 2022. The technology developed by OpenAI is trained to track user questions and provide detailed responses[3]. In the digital age, access to health-related information has expanded beyond traditional sources like doctors, printed materials, and common search engines. This change in how people look for information is especially notable among younger individuals, who make up a large part of chatbot users and are also more likely to seek information about sexually transmitted infections such as HPV. Issues related to sexual health, including HPV infection, may be perceived as sensitive or stigmatizing, leading some individuals to avoid direct consultation with healthcare providers. In this context, the anonymity and accessibility of AI-based chat systems may make them a preferred initial source of information, as similarly observed in studies on AI-generated responses to musculoskeletal conditions[4].

In our study, we aimed to evaluate the accuracy, consistency, and reproducibility of ChatGPT's responses to questions that patients may have about HPV infection, including diagnosis, risk factors, transmission, prevention, vaccination, treatment, management of complications, and screening.

[1]Dokuz Eylul University, School of Medicine, Department of Obstetrics and Gynecology, Division of Gynecologic Oncology – Izmir, Turkey.
[2]Koc University, School of Medicine, Department of Obstetrics and Gynecology, Division of Gynecologic Oncology – Istanbul, Turkey.
*Corresponding author: oguzarslan29@hotmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on September 14, 2025. Accepted on December 19, 2025.
Scientific Editor: José Maria Soares Júnior

METHODS

A total of 123 questions about HPV infection were prepared to ask ChatGPT. Questions were collected from the frequently asked questions pages of the World Health Organization, Pan American Health Organization, Centers for Disease Control and Prevention, United Nations Children's Fund, and the immunize.org website. Questions with the same meaning, unrelated to HPV infection, and unclear meaning were not included in the study. The questions were edited for grammar and spelling. We considered how patients would ask ChatGPT the questions they wanted to learn the answers to, and we intentionally rephrased the questions like they were written in the first person and revised them into colloquial words. The questions were asked to the model in May and July 2024. The questions were categorized into four categories: basic knowledge, diagnosis, treatment, and preventive medicine, to enable domain-specific evaluation of ChatGPT's performance. This classification approach has been previously used in studies assessing ChatGPT responses in various medical fields, allowing comparison across clinically relevant content areas rather than relying solely on overall accuracy metrics[5-7]. ChatGPT version 3.5 was utilized in the study. ChatGPT 3.5 was preferred due to its free and would be easier to access. The decision to choose the free version in this study was influenced by the fact that HPV infection is more common in low- and middle-income countries and the more significant usage of the free version in these regions. Each question was independently entered into ChatGPT twice at different times, and two separate answers were obtained. This method was aimed at evaluating the consistency of ChatGPT's responses to the questions. The model was prompted to respond using the new chat feature when transitioning from one question to another.

We reviewed ChatGPT's answers for both accuracy and consistency. Answers were evaluated for accuracy in four categories: 1. Comprehensive (the answer is correct, and the gyneco-oncologist has no additional comments to add to ChatGPT's response), 2. Correct but inadequate (the answer is correct but incomplete, and the gyneco-oncologist has information to add to ChatGPT's response.), 3. Some correct and some incorrect, 4. Completely incorrect. To reduce subjectivity in accuracy assessment, predefined criteria were applied. Responses were classified as comprehensive if they were factually correct, addressed all key clinical and educational aspects of the question, and required no additional clarification or supplementation by a gyneco-oncologist. Responses were classified as correct but inadequate if they were factually accurate but incomplete, lacking one or more essential elements that would be included in standard patient counseling, such as clarification of risks, limitations, indications, contraindications, or contextual explanations. Given that ChatGPT version 3.5 is limited by a January 2022 knowledge cutoff, this temporal constraint was taken into consideration during the evaluation of its responses. We evaluated ChatGPT's two answers to each question to analyze consistency. Two groups were formed to evaluate the answers: Group 1: Categories 1 and 2, Group 2: Categories 3 and 4. At the end of the scoring process, answers assigned to the same group were considered consistent and reproducible, reflecting rating-level reproducibility rather than semantic equivalence of the response content. Answers were independently evaluated for accuracy and consistency by gyneco-oncologist (OA and BG). At the end of the evaluation, the two reviewers accepted answers if they were consistent across evaluation categories and groups. In cases of disagreement between the two primary reviewers, the response was reviewed by a third board-certified senior gynecologic oncologist (MA) with 39 years of experience in gynecologic oncology, who facilitated a consensus decision. The third reviewer did not independently reassess all responses but was involved only in resolving discordant evaluations. The three reviewers were independent and blind to each other when assessing the questions. Accuracy assessments were informed by current evidence-based clinical guidelines and standard practice recommendations.

Ethics committee approval was not received due to the nature of the study and the fact that the answers provided by ChatGPT were public. Data was analyzed by Statistical Package for the Social Sciences (Version 21.0, IBM SPSS Statistics for Windows; IBM Corp. Armonk, NY, USA). Frequency and ratio values were used for descriptive statistics. The questions were evaluated both within the category and overall. A Kappa analysis was performed to assess the agreement between the two reviewers.

RESULTS

Of the 123 questions prepared for entry into ChatGPT related to HPV infection, 52% were categorized as preventive medicine, 28.5% as basic knowledge, 13.8% as diagnosis, and 5.7% as treatment. Only a subset of representative questions is shown in Table 1.

The model provided "comprehensive" answers to 93.5% of 123 questions, "correct but inadequate" answers to 4.1%, and "some correct, some incorrect" answers to 2.4%. At the end of the evaluation of all answers, no response was detected as "completely incorrect." The distribution of comprehensive answers by category is 100% in the treatment category, followed by basic knowledge at 94.3%, preventive medicine at 93.8%,

and diagnosis at 88.2%. The rate of "correct but inadequate" responses was 5.7 and 4.7% in the basic knowledge and preventive medicine categories, respectively, with no responses as correct but inadequate in the diagnosis and treatment categories. The rate of "some correct, some incorrect" responses was 11.8% in the diagnosis category, 1.6% in the preventive medicine category, and none in the basic knowledge and treatment categories (Table 2).

ChatGPT provided reproducible answers to 120 (97.6%) of the questions. The distribution of reproducible responses by category is as follows: treatment 100%, preventive medicine 98.4%, basic knowledge 97.1%, and diagnosis 94.1% (Table 3).

Reviewers 1 and 2 evaluated ChatGPT's answers and determined that 95 (77.2%) questions were compatible. After the two reviewers' evaluation, the third reviewer evaluated and resolved 28 (22.8%) incompatible questions. In the kappa analysis conducted to test the reliability between the first two reviewers, the kappa value was 0.746, indicating a substantial level of agreement between them[8].

DISCUSSION

Since its launch, ChatGPT has continued to grow in popularity, consistently surprising people with its capabilities. In this study, 52% of the questions submitted to the model were related to preventive medicine. Due to the nature of publicly available HPV-related educational materials, questions focusing on preventive medicine, particularly those addressing vaccination and protection, were more frequently represented in the question set. This distribution likely reflects real-world patient information-seeking behavior, in which prevention-oriented topics constitute a substantial proportion of publicly accessible health information. Our study found that the model's answers to questions about HPV infection were 93.5% comprehensive, 97.6% reproducible, and consistent. In the study by Samaan et al., which evaluated ChatGPT's answers to questions about bariatric surgery, it was shown that the model provided responses that were 86.8% comprehensive and 90.7% reproducible[5]. In Yeo et al.'s study assessing the performance of ChatGPT, the model's comprehensive and correct response rates to questions about cirrhosis were 47.3 and 79.1%, and its comprehensive and correct response rates to questions about hepatocellular cancer were 41.1 and 74%, respectively[6]. In the study by Hermann et al., which analyzed ChatGPT's accuracy regarding cervical cancer, the comprehensiveness rate of the model's answers was determined to be 53.1%[7]. As shown, some results from studies in the current literature that assess ChatGPT's knowledge of medical subjects are consistent with

Table 1. Questions entered into Chat Generative Pre-trained Transformer by category.

Basic knowledge n=35 (28.5)	Preventive medicine n=64 (52)
1. What is human papillomavirus (HPV)? 2. How can HPV be transmitted? 3. Is sexual contact the only way of getting HPV? 4. How long does HPV infection last? 5. Does HPV cause cancer? 6. What types of HPV infection cause cancer? 7. Can all HPV types cause cervical cancer?	1. What is the HPV vaccine? 2. When did the HPV vaccination program start? 3. Can HPV infection be prevented? 4. How can you protect against HPV infection? 5. How are HPV infection and cervical pre-cancer screened for? 6. How can HPV and cervical cancer be prevented? 7. What vaccines are available to prevent HPV? 8. Who recommends the HPV vaccine? 9. What are the HPV vaccines composed of? 10. When should people get the HPV vaccine? 11. How effective is the HPV vaccine? 12. When were the HPV vaccines licensed? 13. How is this HPV vaccine given? 14. What protection is provided by one HPV vaccine dose? 15. Who should get the HPV vaccine? 16. Should individuals be screened before getting HPV vaccinated? 17. Why does my child need HPV vaccination? 18. How long does HPV vaccine protection last? 19. What is the recommended schedule for HPV vaccines? 20. Are the HPV vaccines productive and effective in preventing cervical cancer? 21. Can the HPV vaccine be co-administered with other vaccines? 22. Do the children need consent from their parents or caregivers to be HPV vaccinated? 23. Does the vaccine protect against all types of HPV?
Diagnosis n=17 (13.8)	
1. How is HPV infection diagnosed? 2. Why test for HPV? 3. If my HPV test was positive, what can I tell my partner? 4. Can I ask for an HPV test at any time? 5. If I have HPV, can I ask for my partner to have a test too? 6. Can I get the same type of HPV more than once? 7. How do we test for high-risk HPV?	
Treatment n=7 (5.7)	
1. Can HPV infection be treated? 2. How are genital warts treated? 3. Is there an immune response after a natural HPV infection? 4. What does the "screen-and-treat" approach mean in HPV? 5. How are HPV infection and cervical pre-cancer treated? 6. As I get older, will I be less able to get rid of HPV? 7. In the past, I had cervical cell changes that went away by themselves. Does this mean my immune system is now strong enough to fight HPV?	

Values are presented n (%). HPV: human papillomavirus.

our findings, but some are not. We believe that this difference arises for two reasons. The first reason is that the studies investigate different areas of medicine and various medical subjects. In our study, we selected HPV infection, a common condition that affects all sexually active men and women in the population. In line with this choice, the questions we prepared for ChatGPT focused primarily on topics of broad public concern related to HPV infection, such as transmission, screening, prevention, and vaccination, rather than on technical or exam-related medical information. Cirrhosis, hepatocellular cancer, and bariatric surgery are health problems that affect only a portion of the population and are not as common and frequent as HPV infection. The second reason, we believe, arises from differences in the number of questions used in the studies. The number of questions in our study and Samaan's study was 123 and 151, respectively. Hermann et al. used 64 questions in their study, nearly half the number used in our study. Increasing the number of questions used in studies will allow for a better evaluation of ChatGPT's performance and result in stronger statistical analyses.

Table 2. Grading of responses generated by Chat Generative Pre-trained Transformer to questions related to human papillomavirus infection categorized by question type.

Basic knowledge (n=35)	
Comprehensive	33 (94.3)
Correct but inadequate	2 (5.7)
Some correct and some incorrect	0 (0)
Completely incorrect	0 (0)
Diagnosis (n=17)	
Comprehensive	15 (88.2)
Correct but inadequate	0 (0)
Some correct and some incorrect	2 (11.8)
Completely incorrect	0 (0)
Treatment (n=7)	
Comprehensive	7 (100)
Correct but inadequate	0 (0)
Some correct and some incorrect	0 (0)
Completely incorrect	0 (0)
Preventive medicine (n=64)	
Comprehensive	60 (93.8)
Correct but inadequate	3 (4.7)
Some correct and some incorrect	1 (1.6)
Completely incorrect	0 (0)
Total (n=123)	
Comprehensive	115 (93.5)
Correct but inadequate	5 (4.1)
Some correct and some incorrect	3 (2.4)
Completely incorrect	0 (0)

Values are presented n (%).

Table 3. The proportion of questions with reproducible responses generated by Chat Generative Pre-trained Transformer.

Category of the questions	Reproducible responses
Basic knowledge (n=35)	34 (97.1)
Diagnosis (n=17)	16 (94.1)
Treatment (n=7)	7 (100)
Preventive medicine (n=64)	63 (98.4)
Total (n=123)	120 (97.6)

Values are presented n (%).

HPV is highly prevalent, with reports indicating that 84.6% of women and 91.3% of men will be infected at some point in their lives[9]. A study conducted in the United States reported an HPV prevalence of 40%[10]. Although HPV is very common, public health literacy about it is not at the desired level. Studies have reported that increased health literacy regarding HPV leads to higher rates of Pap smear screening and vaccination for protection against infection[11,12]. The tailored solutions developed by policymakers to improve access to accurate health-related information and health literacy do not reach the entire public promptly and effectively. In some regions, patients have difficulty accessing healthcare providers[13]. We believe that ChatGPT can help bridge this gap. The model will help alleviate patients' concerns at the initial step with its performance in both general and sub-domains (basic knowledge, diagnosis, treatment, and prevention medicine) related to HPV infection. Patients can benefit from AI technology that communicates with them conversationally, as an informative resource. The accessibility of the free version of ChatGPT v3.5 is an important factor contributing to its increased use. We believe that increasing awareness of HPV, particularly among individuals in countries with low socioeconomic development, will positively impact screening and vaccination rates.

Platforms with online access, particularly popular free social media resources such as YouTube and Twitter, are widely used by patients worldwide as sources of health information[14]. Patients seeking information about HPV infection, which ranks first among sexually transmitted diseases, also refer to social media content. The accuracy and reliability of content produced in digital media are unclear. The results of studies examining social media content regarding HPV infection are insufficient and inconsistent with each other. While health

professionals share information on these platforms, individuals with ulterior motives aim to spread false and harmful information. It is easy to access both accurate and inaccurate information, and patients may struggle to assess the accuracy of the information. The spread of misinformation to the public and the rise in false beliefs about HPV infection may complicate the course and management of the disease[15,16]. Since patients cannot discern the accuracy of HPV-related content on social media, ChatGPT could help address this challenge. The advantages of ChatGPT include its high accuracy and repeatability in responses and its use of language that is more comprehensible to patients. Additionally, the model adjusts its responses based on user feedback. We believe that ChatGPT can potentially reduce inequalities in access to health information owing to its advanced features. However, the use of AI-based tools, such as ChatGPT, in healthcare also raises important ethical and medicolegal issues. These include liability for incorrect information, data privacy concerns, and the risk that patients, especially in low-resource settings, may rely on AI-generated advice instead of professional medical care. Therefore, such tools should be implemented with clear safeguards, transparency regarding their limitations, and explicit emphasis on their role as supportive rather than substitutive resources within the healthcare system.

Although ChatGPT possesses impressive features, it can also produce false and inaccurate information. In our study, 2.4% of the model's responses fall into the "some correct, some incorrect" category. Other studies indicate that the model produces incorrect information[6,7]. We believe that the basis of the misinformation generated by the model may be attributed to the lack of transparency regarding the sources of information used in its development, its ongoing development, and the delay in updating with current literature. According to a statement from OpenAI, the model is expected to improve and update its information based on the feedback it receives[17]. Large language models, including ChatGPT, have been shown to generate plausible but incorrect or misleading information, commonly referred to as hallucinations[18]. Although only 2.4% of responses in our study were classified as partially incorrect, even low rates of misinformation may pose clinically meaningful risks in healthcare settings. These findings underscore the importance of expert oversight and emphasize that AI-generated health information should not be used without a suitable clinical context.

This study is the first to evaluate ChatGPT's accuracy and consistency regarding HPV, with independent expert review enhancing its robustness. Given the global prevalence of HPV, the findings have broad applicability. Although response evaluations were informed by current clinical guidelines and expert knowledge, the study did not include a structured, item-level comparison with specific guideline documents. Another limitation of this study is the exclusive evaluation of ChatGPT version 3.5. Although newer versions, such as ChatGPT-4, have demonstrated superior performance, their restricted accessibility may limit real-world use in certain populations. The small number of questions in the treatment category limits the interpretability and generalizability of this finding. no inferential statistical tests or adjustments for multiple comparisons were performed, and category-based differences are therefore descriptive rather than statistically supported. Another limitation of this study is the absence of systematic prompt engineering. This approach was chosen to reflect real-world patient use, in which questions are typically asked in informal, non-optimized language. ChatGPT may produce inaccuracies, and healthcare practices vary; it should serve only as a supplementary resource, with final decisions left to healthcare professionals.

CONCLUSION

Patients often have difficulty accessing reliable information about HPV, as current information sources are limited and unreliable. ChatGPT may positively contribute to the increase in health literacy. These findings are limited to expert-reviewed responses and do not address patient interpretation, behavioral impact, or clinical outcomes. The model may be considered a valuable informational resource for patients when used in conjunction with care provided by healthcare professionals; however, it should be regarded solely as a supplementary tool and should not replace professional medical consultation.

ACKNOWLEDGMENTS

We want to thank Professor Macit Arvas for his valuable contributions.

AUTHORS' CONTRIBUTIONS

OA: Data curation, Formal Analysis, Methodology, Supervision, Writing – original draft, Writing – review & editing. **BG:** Formal Analysis, Supervision, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Serrano B, Brotons M, Bosch FX, Bruni L. Epidemiology and burden of HPV-related disease. Best Pract Res Clin Obstet Gynaecol. 2018;47:14-26. https://doi.org/10.1016/j.bpobgyn.2017.08.006
2. Schiffman M, Doorbar J, Wentzensen N, Sanjosé S, Fakhry C, Monk BJ, et al. Carcinogenic human papillomavirus infection. Nat Rev Dis Primers. 2016;2:16086. https://doi.org/10.1038/nrdp.2016.86
3. Ouyang F, Zheng L, Jiao P. Artificial intelligence in online higher education: a systematic review of empirical research from 2011 to 2020. Educ Inf Technol. 2022;27:7893-925.
4. Keles A, Illeez OG, Erbagci B, Giray E. Artificial intelligence-generated responses to frequently asked questions on coccydynia: evaluating the accuracy and consistency of GPT-4o's performance. Arch Rheumatol. 2025;40(1):63-71. https://doi.org/10.46497/ArchRheumatol.2025.10966
5. Samaan JS, Yeo YH, Rajeev N, Hawley L, Abel S, Ng WH, et al. Assessing the accuracy of responses by the language model ChatGPT to questions regarding bariatric surgery. Obes Surg. 2023;33(6):1790-6. https://doi.org/10.1007/s11695-023-06603-5
6. Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol. 2023;29(3):721-32. https://doi.org/10.3350/cmh.2023.0089
7. Hermann CE, Patel JM, Boyd L, Growdon WB, Aviki E, Stasenko M. Let's chat about cervical cancer: assessing the accuracy of ChatGPT responses to cervical cancer questions. Gynecol Oncol. 2023;179:164-8. https://doi.org/10.1016/j.ygyno.2023.11.008
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74. PMID: 843571
9. Chesson HW, Dunne EF, Hariri S, Markowitz LE. The estimated lifetime probability of acquiring human papillomavirus in the United States. Sex Transm Dis. 2014;41(11):660-4. https://doi.org/10.1097/OLQ.0000000000000193
10. Lewis RM, Laprise JF, Gargano JW, Unger ER, Querec TD, Chesson HW, et al. Estimated prevalence and incidence of disease-associated human papillomavirus types among 15- to 59-year-olds in the United States. Sex Transm Dis. 2021;48(4):273-7. https://doi.org/10.1097/OLQ.0000000000001356
11. Thompson EL, Wheldon CW, Vamos CA, Griner SB, Daley EM. How is health literacy related to pap testing among US women? J Cancer Educ. 2019;34(4):789-95. https://doi.org/10.1007/s13187-018-1375-z
12. Lee HY, Lee J, Henning-Smith C, Choi J. HPV literacy and its link to initiation and completion of HPV vaccine among young adults in Minnesota. Public Health. 2017;152:172-8. https://doi.org/10.1016/j.puhe.2017.08.002
13. Nutbeam D, McGill B, Premkumar P. Improving health literacy in community populations: a review of progress. Health Promot Int. 2018;33(5):901-11. https://doi.org/10.1093/heapro/dax015
14. Lin W-Y, Zhang X, Song H, Omori K. Health information seeking in the web 2.0 age: trust in social media, uncertainty reduction, and self-disclosure. Comput Human Behav. 2016;56:289-94.
15. Ortiz RR, Smith A, Coyne-Beasley T. A systematic literature review to examine the potential for social media to impact HPV vaccine uptake and awareness, knowledge, and attitudes about HPV and HPV vaccination. Hum Vaccin Immunother. 2019;15(7-8):1465-75. https://doi.org/10.1080/21645515.2019.1581543
16. Zimet GD, Rosberger Z, Fisher WA, Perez S, Stupiansky NW. Beliefs, behaviors and HPV vaccine: correcting the myths and the misinformation. Prev Med. 2013;57(5):414-8. https://doi.org/10.1016/j.ypmed.2013.05.013
17. OpenAI. ChatGPT: optimizing language models for dialogue [Internet]. OpenAI; 2023. [Cited on 2024 Apr 28]. Available from: https://openai.com/blog/chatgpt/
18. Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y et al. Survey of hallucination in natural language generation. ACM Comput Surv. 2023;55:1-38.

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251795

Impact of maternal nutrition and gestational weight gain on perinatal outcomes in intrahepatic cholestasis of pregnancy

Esra Keles[1]*, Leyla Kaya[2], Yasemin Tugba Ogunc[3], Zahide Kaya[4], Kurşad Nuri Baydili[5], Pınar Kumru[6]

SUMMARY

OBJECTIVE: The aim of this study was to investigate the relationship between intrahepatic cholestasis of pregnancy and maternal nutritional intake, and to evaluate the impact of maternal nutritional status on maternal and fetal outcomes.

METHODS: This prospective case-control study enrolled 38 women with intrahepatic cholestasis of pregnancy and 76 age- and body mass index-matched controls between 2022 and 2023. Dietary intake was assessed using 24-h recall interviews and a 60-item food frequency questionnaire. Nutrient adequacy was calculated relative to age- and pregnancy-specific Dietary Reference Intakes. Maternal anthropometrics and gestational weight gain were recorded, alongside maternal and neonatal outcomes.

RESULTS: Women with intrahepatic cholestasis of pregnancy had significantly lower energy intake (p=0.004) and gestational weight gain (p<0.001). Nutrient adequacy analyses indicated comprehensive deficiencies, with significantly lower adequacy for energy, protein, fiber, folate, magnesium, potassium, and iron (all p<0.01). Intrahepatic cholestasis of pregnancy was associated with earlier delivery (p<0.001), lower neonatal birth weight (p<0.001), and higher neonatal intensive care unit admission rates (p=0.007). Within the intrahepatic cholestasis of pregnancy group, maternal fasting bile acid levels inversely correlated with gestational age at delivery (ρ=-0.408, p=0.011). Gestational weight gain positively correlated with neonatal birth weight (ρ=0.325, p=0.046).

CONCLUSION: Intrahepatic cholestasis of pregnancy is associated with significant maternal nutritional inadequacy and insufficient gestational weight gain. Optimizing maternal nutritional status represents a potentially modifiable pathway for targeted intervention in the comprehensive management of intrahepatic cholestasis of pregnancy.

KEYWORDS: Cholestasis, intrahepatic. Nutritional status. Premature birth. Birth weight. Weight gain. Diet.

INTRODUCTION

Intrahepatic cholestasis of pregnancy (ICP) is recognized as the most frequent reversible liver disorder specific to gestation, affecting approximately 0.1–2% of pregnant women[1]. Clinically, the condition is defined by maternal pruritus and pathological elevation of serum bile acid (BA) concentrations, typically resolving rapidly following delivery. While the maternal prognosis is generally favorable, ICP presents considerable risk to fetal well-being.

The etiology of ICP involves complex interplay among genetic susceptibility, hormonal factors, and environmental influences[2]. Pathophysiologically, impaired hepatocellular secretion of BAs leads to their accumulation in the maternal circulation. These elevated BA concentrations are strongly associated with increased risk of adverse perinatal outcomes, including spontaneous preterm birth, meconium staining, and sudden intrauterine death, especially when BA levels exceed 100 µmol/L[3]. Furthermore, neonates born to mothers with ICP frequently exhibit indicators of fetal growth restriction[4].

Cholestasis imposes a dual challenge regarding maternal nutritional status. First, impaired bile flow directly impedes the formation of micelles in the intestine. This leads to malabsorption of dietary lipids and fat-soluble vitamins[5]. Second, clinical symptoms such as severe pruritus and potential nausea may substantially decrease maternal appetite and overall caloric

[1]University of Health Sciences, Kartal Dr. Lütfi Kırdar City Hospital, Department of Gynecologic Oncology – Istanbul, Turkey.
[2]University of Health Sciences, Faculty of Health Sciences, Department of Midwifery – Istanbul, Turkey.
[3]University of Health Sciences, Hamidiye Faculty of Health Sciences, Department of Nutrition and Dietetics – Istanbul, Turkey.
[4]Uskudar State Hospital, Internal Medicine Clinic – Istanbul, Turkey.
[5]University of Health Sciences Turkey, Hamidiye Faculty of Medicine, Department of Biostatistics – Istanbul, Turkey.
[6]University of Health Sciences, Zeynep Kamil Women and Children's Disease Training and Research Hospital, Department of Obstetrics and Gynecology – Istanbul, Turkey.
*Corresponding author: dresrakeles@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on October 19, 2025. Accepted on January 24, 2026.
Scientific Editor: José Maria Soares Júnior

intake[6]. Suboptimal nutrient and energy consumption, combined with malabsorption, can rapidly deplete maternal reserves, potentially contributing to inadequate gestational weight gain (GWG)[7]. Although optimizing diet is widely accepted as beneficial for overall pregnancy health, current guidelines often lack specific dietary recommendations for women with ICP due to limited evidence.

The primary objective of this study was to compare the energy and nutrient intake, dietary adequacy relative to established Dietary Reference Intakes (DRIs), and anthropometric indices, including GWG, between pregnant women diagnosed with ICP and healthy controls. The secondary objective was to evaluate the association between maternal nutritional parameters and adverse maternal and neonatal outcomes.

METHODS

Study design and population

This prospective case-control study was conducted in the Department of Obstetrics and Gynecology at a tertiary referral hospital between June 2023 and June 2024. The cohort comprised 114 pregnant women: 38 diagnosed with ICP (cases) and 76 age- and body mass index-matched healthy pregnant women (controls). Ethical approval was granted by the Hospital's Ethics Committee (Approval 88, May 24, 2023), following the Declaration of Helsinki tenets. Diagnosis of ICP was based on clinical presentation of pruritus and elevated serum BA concentrations (>10 μmol/L). Patients with other underlying hepatic disorders or systemic diseases that could mimic cholestasis were excluded.

Data collection and anthropometric measurements

Maternal sociodemographic factors, lifestyle habits, and obstetric history were collected using a standardized questionnaire. All anthropometric assessments were conducted by a dietitian following standardized techniques. Weight and height were recorded to the nearest 0.1 kg and 1 cm, respectively. Body mass index was derived and classified per World Health Organization guidelines[8]. GWG was calculated as the difference between self-reported pre-pregnancy weight and weight measured at delivery. Waist circumference was assessed at the mid-abdominal level during mid-exhalation. Hip circumference was measured at the maximal posterior protrusion of the buttocks.

Nutritional assessment

Dietary intake assessment was conducted by registered dietitians using validated self-report methods. A combination of 24-h dietary recall and a food frequency questionnaire was employed. The 24-h dietary recall was conducted through a face-to-face method with a Food and Dish Photograph Catalog serving as a retrospective reminder technique[9]. The Food Frequency Questionnaire comprised 60 items, categorized into food groups. For each item, participants reported their consumption frequency over the preceding 3-month period and typical portion size. Daily nutrient intakes were calculated using the Turkish food composition database (BEBIS, Version 8.0)[10]. Nutrient intakes were compared against DRIs established for the Turkish population, adjusted for pregnancy requirements[11].

Dietary diversity assessment

Dietary diversity was assessed using a 24-h recall and quantified via the Women's Dietary Diversity Score based on Food and Agriculture Organization guidelines[12]. Dietary intake was categorized into nine food groups. A score of one point was assigned for each food group if the consumed amount was ≥15 g.

Maternal and neonatal outcomes

Maternal outcomes included gestational age at delivery, mode of delivery, maternal length of hospital stay, and fasting serum BA concentrations. Neonatal outcomes included birth weight, length, head circumference, Apgar scores at 1 and 5 min, need for neonatal intensive care unit admission, and requirement for respiratory support.

Statistical analysis

Statistical analyses were performed using Statistical Package for the Social Sciences version 28.0. Normality of continuous variables was assessed using the Kolmogorov-Smirnov test. Continuous variables were summarized as mean±standard deviation for normally distributed data or median (minimum–maximum) for skewed data. Categorical variables were expressed as frequencies and percentages. Group comparisons were conducted using Student's t-test or the Mann-Whitney U test for continuous variables, and the chi-square or Fisher's exact test for categorical variables. Correlations were examined using Spearman's rho (ρ). Statistical significance was set at $p<0.05$.

RESULTS

Study population characteristics

The study included 114 pregnant women, with 38 in the ICP group and 76 in the control group. No significant differences were observed between groups regarding maternal education, occupation, family structure, pregnancy planning status,

smoking habits, or consumption of coffee and tea. A history of cholestasis in a previous pregnancy was significantly more common in the ICP group (p<0.001). Furthermore, the ICP group exhibited a higher rate of co-existing gestational diabetes compared to controls (47.4 vs. 15.8%; p<0.001).

Maternal characteristics and anthropometric measurements

Maternal age did not differ significantly between groups (ICP: 30 [22–41] years; Control: 28 [18–38] years; p=0.068). Pre-pregnancy body mass index (25.8±5.6 vs. 25.3±4.7 kg/m²; p=0.759) was comparable between groups. GWG was significantly lower in women with ICP compared to controls (9.3±6.5 vs. 12.9±6.6 kg, respectively; p<0.001) (Table 1).

Pregnancy and delivery outcomes

Maternal length of hospital stay was significantly prolonged in the ICP group (median 10 days [range: 3–18] vs. 2 days [range: 1–5], p<0.001). Women with ICP delivered at significantly earlier gestational ages compared to controls (median 37 weeks [range: 32.3–39.6] vs. 39 weeks [range: 35–41.2], p<0.001) (Table 1).

Neonatal outcomes

Neonatal anthropometric measurements were lower in infants born to mothers with ICP. Birth weight was significantly reduced (2,795.5±575.3 vs. 3,272.6±483.0 g, p<0.001), as were birth length (48.3±2.6 vs. 50.2±2.1 cm, p<0.001) and head circumference (median 34 cm [range: 30–37] vs. 35 cm [range: 31–39], p=0.009). Apgar scores were significantly lower in the ICP group at both 1 min (median 8 [range: 4–8] vs. 8 [range: 6–8], p=0.021) and 5 min (median 9 [range: 6–9] vs. 9 [range: 7–9], p=0.002). Neonates in the ICP group required intensive care admission more frequently than controls (56 vs. 27%, p=0.007) and had higher rates of respiratory support requirements (56 vs. 27%, p=0.007) (Table 1).

Nutritional intake assessment

Women with ICP demonstrated significantly lower total energy intake compared to controls (median 1,850.7 kcal/day [range: 1,287.1–2,604.2] vs. 2,055.6 kcal/day [range: 1,026.3–3,758.2], p=0.004). Macronutrient analysis revealed significantly reduced intake of protein (64.4 vs. 70.0 g, p=0.023), carbohydrates (199.1 vs. 217.5 g, p=0.027), and total fat (83.0 vs. 95.1 g, p=0.010) in the ICP group.

Fiber intake was significantly lower in women with ICP, with total fiber intake of 39.1 g compared to 47.5 g in controls (p=0.001). both soluble fiber (6.7 vs. 7.9 g, p=0.007) and insoluble fiber (12.4 vs. 14.9 g, p=0.001) were significantly reduced in the ICP group. multiple micronutrient deficiencies were identified in the ICP group. vitamin A intake was significantly lower (median 1,011.5 vs. 1,327.9 µg, p=0.001), as was folate intake from dietary sources (321.1 vs. 381.1 µg, p<0.001), and vitamin C intake (92.7 vs. 112.9 mg, p=0.001). mineral intake analysis revealed significant deficiencies in potassium (2,468.9 vs. 2,926.3 mg, p=0.002), magnesium (261.5 vs. 305.9 mg, p=0.002), and iron (9.2 vs. 10.7 mg, p=0.002) (Table 2).

Table 1. Maternal and neonatal characteristics of control and cholestasis groups.

Variable	Control (n=76)	Cholestasis (n=38)	p value
Maternal age (years)	28 (18–38)	30 (22–41)	0.068
Gestational age at delivery (weeks)	39 (35–41.2)	37 (32.3–39.6)	<0.001
Pre-pregnancy BMI (kg/m²)	25.3±4.7	25.8±5.6	0.656
Gestational weight gain (kg)	12.9±6.6	9.3±6.5	<0.001
Waist circumference (cm)	114 (98–130)	109 (92–130)	0.003
Hip circumference (cm)	113 (96–126)	110 (94–132)	0.004
Birth weight (g)	3,272.6±483.0	2,795.5±575.3	<0.001
Birth length (cm)	50.2±2.1	48.3±2.6	<0.001
Head circumference (cm)	35 (31–39)	34 (30–37)	0.009
Apgar score, 1 min	8 (6–8)	8 (4–8)	0.021
Apgar score, 5 min	9 (7–9)	9 (6–9)	0.002
Maternal length of hospital stay (days)	2 (1–5)	10 (3–18)	<0.001
NICU admission, n (%)	11 (27)	14 (56)	0.007
Respiratory support, n (%)	11 (27)	14 (56)	0.007

Values are mean±SD or median (min–max). BMI: body mass index; NICU: neonatal intensive care unit.

Nutrient adequacy compared to dietary reference intakes

When nutrient intakes were compared to established age- and pregnancy-specific DRIs, women with ICP demonstrated significantly lower adequacy percentages across multiple nutrients. Energy adequacy was 82.0% in the ICP group compared to 91.3% in controls (p=0.006). Protein adequacy was 78.6 vs. 87.9% (p=0.011), and fiber adequacy was significantly reduced at 156.4 vs. 190.1% (p=0.001). Micronutrient adequacy patterns showed concerning deficiencies in the ICP group. Folate adequacy was 53.5% compared to 63.5% in controls (p<0.001), vitamin C adequacy was 88.3 vs. 107.5% (p=0.001), and iron adequacy was 57.2 vs. 66.7% (p=0.002) (Table 2).

Correlation analysis within the intrahepatic cholestasis of pregnancy group

Analysis of associations within the ICP cohort revealed that higher maternal fasting BA concentrations were negatively correlated with gestational age at delivery (ρ=-0.408, p=0.011). Maternal nutritional and anthropometric parameters demonstrated significant positive correlations with fetal growth markers. Neonatal birth weight was positively correlated with maternal GWG (ρ=0.325, p=0.046), waist circumference (ρ=0.474, p=0.003), and hip circumference (ρ=0.384, p=0.017) (Table 3).

DISCUSSION

The results of this prospective case-control study demonstrated a significant association between ICP and compromised maternal nutritional status. Women with ICP exhibited significantly reduced GWG and inadequacies across multiple macro- and micronutrient categories. These nutritional deficits appear to contribute to adverse maternal and neonatal outcomes, including insufficient GWG, preterm delivery, low birth weight, and increased neonatal morbidity.

The observed reduction in median energy intake (approximately 200 kcal/day) and the significantly lower GWG in the ICP group represent a clinically meaningful difference that falls short of recommended pregnancy requirements. The mechanism underlying reduced energy intake in ICP may relate to pregnancy-associated nausea, early satiety due to hepatomegaly, or psychological factors related to chronic pruritus and sleep disturbance[13]. Suboptimal GWG is a recognized independent risk factor for fetal growth restriction[14].

Beyond caloric restriction, the findings of this study revealed a broad pattern of nutritional deficiencies, including protein, polyunsaturated fatty acids, and fat-soluble vitamins. This pattern is consistent with the impaired enterohepatic circulation caused by bile flow obstruction[15]. Bile acids are essential for fat emulsification and absorption. Therefore, their reduced availability leads to steatorrhea and malabsorption of fat-soluble

Table 3. Correlations between maternal parameters and outcomes in the cholestasis group.

Variable	Gestational age (weeks) ρ (p)	Birth weight (g) ρ (p)
Maternal weight gain (kg)	0.236 (0.155)	0.325 (0.046)
Waist circumference (cm)	0.242 (0.144)	0.474 (0.003)
Hip circumference (cm)	0.195 (0.240)	0.384 (0.017)
Fasting bile acids (µmol/L)	-0.408 (0.011)	-0.230 (0.166)

Spearman's rho correlation coefficients shown. p<0.05 considered significant.

Table 2. Daily nutrient intakes compared with dietary reference intakes.

Nutrient	Control (n=76)	Cholestasis (n=38)	p value (intake)	p value (%DRI)
Energy (kcal)	2,055.6 (1,026–3,758)	1,850.7 (1,287–2,604)	0.004	0.006
Protein (g)	70.0 (41.8–117.9)	64.4 (44.7–101.5)	0.023	0.011
Carbohydrate (g)	217.5 (63.2–433)	199.1 (93.1–313.3)	0.027	0.027
Fat (g)	95.1 (62.7–187.1)	83.0 (54.5–124.3)	0.010	-
Fiber (g)	47.5 (26.2–95.2)	39.1 (27.1–65.2)	0.001	0.001
Vitamin A (µg)	1,327.9 (502–3,214)	1,011.5 (563–3,507)	0.001	0.001
Folate (µg)*	381.1 (180–725)	321.1 (219–474)	<0.001	<0.001
Vitamin C (mg)	112.9 (41–346)	92.7 (47–163)	0.001	0.001
Potassium (mg)	2,926.3 (1,638–4,512)	2,468.9 (1,722–3,775)	0.002	0.002
Magnesium (mg)	305.9 (190–474)	261.5 (176–382)	0.002	0.002
Iron (mg)	10.7 (6.4–17.4)	9.2 (5.9–14.5)	0.002	0.002

Values are median (min–max). DRI: Dietary Reference Intakes. *Folate from diet only; supplemental folic acid not included.

nutrients[16]. Furthermore, the observed deficits in essential water-soluble vitamins and minerals suggest a general reduction in dietary quality and intake.

The negative correlation observed between maternal fasting BA concentrations and gestational age at delivery reinforces the critical role of BA severity in determining the timing of delivery[3]. However, the present study further demonstrates that maternal nutritional status acts as a significant modifier of fetal growth, independent of BA concentration. The reduced neonatal birth weight, length, and head circumference confirm the association between ICP and fetal growth restriction, consistent with findings from previous meta-analyses[4]. Importantly, the positive correlation of neonatal birth weight with maternal GWG and other anthropometric reserves suggests that, even in the presence of pathological BA elevation, greater maternal reserves may provide a crucial protective buffer.

The finding that gestational diabetes was prevalent in the ICP group (47.4 vs. 15.8%) suggests a shared metabolic dysfunction, likely involving dysregulated lipid and glucose homeostasis that is exacerbated by elevated BAs[17]. This comorbidity adds complexity to nutritional management. Dietary optimization must address both caloric adequacy for GWG and carbohydrate control for diabetes management.

The findings of the present study indicate that clinical management protocols for ICP should be broadened beyond current approaches. While existing protocols primarily focus on ursodeoxycholic acid and timely delivery planning, they should also incorporate systematic nutritional screening and targeted intervention. Given the comprehensive inadequacies observed, particularly in nutrients such as folate, iron, and protein, specialized dietary assessment and counseling are recommended for all women diagnosed with ICP. Optimizing maternal nutritional reserves represents a modifiable target for improving fetal growth and mitigating perinatal morbidity.

Limitations

Several limitations should be acknowledged. The case-control design precludes causal inference. Thus, associations between dietary intake, GWG, and perinatal outcomes should be interpreted cautiously, as directionality cannot be established, and nutritional deficiencies may represent both causes and consequences of ICP. Dietary intake was self-reported using validated 24-h dietary recalls and food frequency questionnaires, which are subject to recall and reporting bias, although assessments were conducted by trained dietitians using standardized protocols. The cross-sectional assessment of nutritional status limits evaluation of temporal relationships across gestation, necessitating longitudinal studies with serial assessments. Finally, the single-center setting may limit generalizability, and the high prevalence of coexisting gestational diabetes mellitus represents a relevant confounder that should be addressed in future studies.

CONCLUSION

ICP is associated with maternal nutritional compromise characterized by significantly reduced energy and micronutrient intake, and insufficient GWG. These nutritional inadequacies are significantly associated with adverse perinatal outcomes, including earlier gestational age at delivery, lower birth weight, and increased neonatal morbidity requiring neonatal intensive care. The positive association between maternal GWG and neonatal birth weight suggests that optimizing maternal nutritional status may provide a valuable strategy to mitigate the neonatal consequences of ICP. Integration of nutritional assessment and individualized dietary counselling into the medical management of ICP is recommended.

ETHICS COMMITTEE APPROVAL

The database management was in accordance with privacy legislation, and the presented study was in accordance with the ethical principles of the Declaration of Helsinki. Ethical approval for this study was obtained by the Research Ethics Committee (Approval No: 88; Date: 24 May 2023). Written and verbal informed consent was obtained from all study participants.

AUTHORS' CONTRIBUTIONS

EK: Conceptualization, Data curation, Supervision, Formal Analysis, Writing – original draft. **LK:** Data curation, Writing – original draft. **YTO:** Formal Analysis, Writing – original draft. **ZK:** Data curation, Writing – original draft. **KNB:** Formal Analysis, Writing – original draft. **PK:** Writing – original draft.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Dajti E, Tripodi V, Hu Y, Estiù MC, Shan D, Mazzella G, et al. Intrahepatic cholestasis of pregnancy. Nat Rev Dis Primers. 2025;11(1):51. https://doi.org/10.1038/s41572-025-00633-2
2. Yang Z, Yao M, Zhang C, Hu X, Zhong Y, Xu X, et al. Application of metabolomics in intrahepatic cholestasis of pregnancy: a systematic review. Eur J Med Res. 2022;27(1):178. https://doi.org/10.1186/s40001-022-00802-z
3. Ovadia C, Seed PT, Sklavounos A, Geenes V, Ilio C, Chambers J, et al. Association of adverse perinatal outcomes of intrahepatic cholestasis of pregnancy with biochemical markers: results of aggregate and individual patient data meta-analyses. Lancet. 2019;393(10174):899-909. https://doi.org/10.1016/S0140-6736(18)31877-4
4. Li L, Chen YH, Yang YY, Cong L. Effect of ıntrahepatic cholestasis of pregnancy on neonatal birth weight: a meta-analysis. J Clin Res Pediatr Endocrinol. 2018;10(1):38-43. https://doi.org/10.4274/jcrpe.4930
5. Sarker M, Getrajdman C, Warren L, Ferrara L. Severe fat-soluble vitamin deficiency suspected secondary to intrahepatic cholestasis of pregnancy: a case report. Case Rep Womens Health. 2022;35:e00430. https://doi.org/10.1016/j.crwh.2022.e00430
6. Jasak K, Gajzlerska-Majewska W, Jabiry-Zieniewicz Z, Litwińska-Korcz E, Litwińska M, Ludwin A, et al. Intrahepatic cholestasis of pregnancy: diagnosis, management, and future directions-a review of the literature. Diagnostics (Basel). 2025;15(16):2002. https://doi.org/10.3390/diagnostics15162002
7. Mishra KG, Bhatia V, Nayak R. Maternal nutrition and ınadequate gestational weight gain in relation to birth weight: results from a prospective cohort study in India. Clin Nutr Res. 2020;9(3):213-22. https://doi.org/10.7762/cnr.2020.9.3.213
8. World Health Organization. A healthy lifestyle - WHO recommendations. Geneva: World Health Organization; 2010.
9. Rakıcıoglu N, Tek N, Ayaz A, Pekcan G. Photograph catalog of food and dishes: portion sizes and amounts. 6th ed. Ankara: Merdiven Reklam Tanıtım; 2012.
10. BEBIS. Computer aided nutrition program, nutrition ınformation systems package program (BEBIS) Version: 8.1. Istanbul: Mavi Elma Yazilim; 2017. Available from: https://bebis.com.tr/anasayfa
11. Turkey Ministry of Health. Turkey dietary guidelines (TUBER)-2022. 2022. Available from: https://hsgm.saglik.gov.tr/depo/birimler/saglikli-beslenme-ve-hareketli-hayat-db/Dokumanlar/Rehberler/Turkiye_Beslenme_Rehber_TUBER_2022_min.pdf
12. FAO. Guidelines for measuring household and individual dietary diversity. Rome: Food and Agriculture Organization of the United Nations; 2007.
13. Geenes V, Chappell LC, Seed PT, Steer PJ, Knight M, Williamson C. Association of severe intrahepatic cholestasis of pregnancy with adverse pregnancy outcomes: a prospective population-based case-control study. Hepatology. 2014;59(4):1482-91. https://doi.org/10.1002/hep.26617
14. Perumal N, Wang D, Darling AM, Liu E, Wang M, Ahmed T, et al. Suboptimal gestational weight gain and neonatal outcomes in low and middle income countries: individual participant data meta-analysis. BMJ. 2023;382:e072249. https://doi.org/10.1136/bmj-2022-072249
15. Schmidt DR, Holmstrom SR, Fon Tacer K, Bookout AL, Kliewer SA, Mangelsdorf DJ. Regulation of bile acid synthesis by fat-soluble vitamins A and D. J Biol Chem. 2010;285(19):14486-94. https://doi.org/10.1074/jbc.M110.116004
16. Majsterek M, Wierzchowska-Opoka M, Makosz I, Kreczyńska L, Kimber-Trojnar Ż, Leszczyńska-Gorzelak B. Bile acids in ıntrahepatic cholestasis of pregnancy. Diagnostics (Basel). 2022;12(11):2746. https://doi.org/10.3390/diagnostics12112746
17. Avsar HA, Atlıhan U, Ata C, Erkılınc S. Intrahepatic cholestasis of pregnancy and its association with preeclampsia and gestational diabetes: a retrospective analysis. Arch Gynecol Obstet. 2024;310(1):221-7. https://doi.org/10.1007/s00404-024-07507-0

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251801

Ciprofol provides safer sedation than propofol in minor gynecologic surgery: evidence from a randomized controlled trial

Wenbo Huang[1†], Qingtong Chen[1†], Dashuo Peng[1†], Yanhui Li[2], Wenyang Liu[1], Sijin Peng[1], Guiling Xie[1*]

SUMMARY

OBJECTIVE: The quality of sedation critically affects the success and safety of minor gynecologic surgeries. Ciprofol, a novel analog of propofol, offers rapid onset and recovery with a potentially improved safety profile. This study compared the efficacy and safety of ciprofol and propofol for sedation during minor gynecologic surgery.

METHODS: In this single-center, randomized, double-blind trial, 342 patients scheduled for minor gynecologic procedures were assigned to receive either ciprofol or propofol for anesthesia induction and maintenance. The primary outcomes were anesthesia onset time and surgical success rate. Secondary outcomes included recovery time, hemodynamic stability, body movement, injection pain, satisfaction ratings, and adverse events such as respiratory depression and hypotension.

RESULTS: Both agents achieved a 100% anesthesia success rate, with comparable onset times (ciprofol: 37.2±10.4 s vs. propofol: 38.7±11.2 s; p=0.214). Ciprofol required a fivefold lower total dose (68.5±31.7 mg vs. 359.6±164.6 mg; p<0.001) and demonstrated superior safety, including fewer body movements (2.4 vs. 15.3%), absence of injection pain (0 vs. 36.5%), and reduced hypotension (10.1 vs. 41.2%; all p<0.001). Satisfaction was higher among surgeons and patients receiving ciprofol.

CONCLUSION: Ciprofol provides anesthesia efficacy equivalent to propofol with significantly enhanced safety and tolerance, supporting its use as a preferred agent for outpatient gynecologic surgery.

KEYWORDS: Ciprofol. Propofol. Gynecologic surgery. Sedation.

INTRODUCTION

In recent years, the pursuit of comfort-oriented medical care has led to the growing adoption of painless diagnostic and therapeutic procedures. Minor gynecologic surgery, characterized by minimal trauma and short operative duration, has become increasingly common. However, the severe pain associated with these procedures often provokes anxiety and fear, which may complicate surgery and necessitate effective anesthetic intervention. Common anesthetic methods for minor gynecologic surgery include local anesthesia, regional block, spinal anesthesia, and general anesthesia. Among these, propofol combined with opioids remains the most widely used regimen owing to its pharmacological advantages, such as rapid onset, predictable metabolism, and strong controllability[1]. Despite these benefits, propofol is associated with notable adverse effects, including circulatory suppression, respiratory depression, injection pain, and propofol infusion syndrome. Therefore, the search for alternative anesthetics with improved safety profiles has become a priority.

Ciprofol, a novel 2,6-disubstituted phenol derivative, is a structural analog of propofol that exhibits stronger binding affinity to the γ-aminobutyric acid A ($GABA_A$) receptor. It has been developed for procedural sedation, induction and maintenance of general anesthesia, and intensive care unit (ICU) sedation. Previous studies have demonstrated that ciprofol at doses of 0.4–0.6 mg/kg produces comparable sedation and recovery profiles to propofol at 1.5–2.5 mg/kg, indicating a wider safety margin[2]. Other investigations have shown that ciprofol possesses a potency ratio of approximately 1/4 to 1/5 compared with propofol and is associated with a lower incidence of respiratory-related adverse events (AEs) and injection pain[3]. Similarly, Luo et al. reported that ciprofol exhibited equivalent efficacy to propofol at one-fifth the dose, with superior safety reflected in fewer drug-related

[1]Heyuan People's Hospital, Department of Anesthesiology – Heyuan, China.
[2]Heyuan People's Hospital, Department of Gynaecology – Heyuan, China.
*Corresponding author: Guiling_Xie@unc-edu.net

Conflicts of interest: the authors declare there is no conflicts of interest. Funding: this study was supported by the Guangdong Medical Research Foundation (Grant Number: A2022372).
Received on August 11, 2025. Accepted December on 22, 2025.
Scientific Editor: Roseli Nomura

adverse reactions and reduced lipid intake during fiberoptic bronchoscopy and ICU sedation[4]. Clinical trials have further confirmed that ciprofol is well tolerated during the induction and maintenance of general anesthesia, resulting in decreased hemodynamic fluctuations, fewer AEs, and significantly alleviated injection pain[5].

Although recent studies have evaluated ciprofol in endoscopy, bronchoscopy, and general anesthesia, clinical evidence specifically supporting its use in minor gynecologic surgery remains limited. This surgical setting requires rapid anesthetic onset, stable hemodynamics, minimal respiratory depression, and a high level of patient comfort—areas in which propofol, despite its widespread use, is often constrained by adverse effects such as hypotension, respiratory suppression, and injection pain. Given ciprofol's higher $GABA_A$ receptor affinity, wider safety margin, and markedly lower incidence of vascular irritation, it may offer significant advantages over propofol. However, direct comparative data in outpatient gynecologic procedures are scarce.

Therefore, we designed a single-center, randomized, double-blind, active-controlled, non-inferiority trial to compare the efficacy and safety of ciprofol versus propofol in patients undergoing minor gynecologic surgery, with the aim of providing high-quality evidence to guide safer and more effective anesthetic selection in outpatient gynecologic practice.

METHODS

Patient selection and eligibility criteria

This study enrolled female patients scheduled for minor gynecologic surgery under sedation. Eligible participants were between 18 and 65 years of age and classified as American Society of Anesthesiologists (ASA) physical status I or II. Physiological stability was ensured by the following inclusion criteria: body mass index (BMI) 18–30 kg/m^2, respiratory rate (RR) 10–24 breaths/min, pulse oxygen saturation (SpO_2) ≥95%, systolic blood pressure (SBP) 85–140 mmHg, diastolic blood pressure (DBP) 50–90 mmHg, and HR 50–100 beats/min.

Patients were excluded if they had a history of alcohol abuse, a known allergy to general anesthetics, or long-term use of sedative-hypnotic or anti-anxiety medications. Those with unregulated or malignant hypertension, severe ischemic heart disease, or significant hepatic or renal dysfunction were also excluded. Pregnant or lactating women, patients with recent respiratory infections, and those with neurological or psychiatric disorders impairing communication were ineligible.

Randomization and blinding

This study adopted a randomized, double-blind design. Participants, the data recorder, and the data analyst were blinded to group assignments. Randomization was performed in a 1:1 ratio using a computer-generated sequence created by an independent researcher in Excel 2010 (Microsoft Office). Group allocation—ciprofol group (Group C) or propofol group (Group P)—was concealed in sequentially numbered envelopes opened by nurses who were not involved in patient management or data analysis.

Study design and procedure

This was a single-center, randomized, parallel-group, active-controlled, non-inferiority trial conducted at Heyuan People's Hospital to compare the efficacy and safety of ciprofol versus propofol in patients undergoing minor gynecologic surgery. A total of 342 patients were randomized equally to the ciprofol or propofol group. All participants fasted for at least 8 h and refrained from clear liquids for at least 2 h before anesthesia induction. Each patient received an infusion of 300–500 mL of compound sodium chloride solution within 1 h before administration of the study drugs. Baseline parameters, including age, weight, height, and ASA grade, were recorded.

During surgery, continuous monitoring of HR, RR, SpO_2, SBP, DBP, mean arterial pressure (MAP), and electrocardiogram (ECG) was performed. Oxygen was administered via mask at 6 L/min for at least 3 min before induction and maintained until full recovery.

The initial induction dose of ciprofol 0.4 mg/kg was selected based on its equivalence to propofol 2.0 mg/kg, consistent with previous studies[2-5]. Both drugs were administered intravenously within 30 s. Sedation depth was assessed using the Modified Observer's Alertness/Sedation (MOAA/S) scale, and surgery commenced when the MOAA/S score reached ≤2. If adequate sedation was not achieved within 2 min, an additional dose of ciprofol 0.2 mg/kg or propofol 1.0 mg/kg was administered intravenously over 10 s. Additional doses were allowed up to five times within 15 min; if sedation remained inadequate, crossover therapy with propofol was initiated.

Efficacy assessments

The primary efficacy endpoint was the onset time of anesthesia, defined as the interval from the first administration of the study drug to the achievement of MOAA/S ≤2.

Secondary efficacy outcomes included: (1) anesthesia success rate, defined as the completion of surgery without the need for an alternative sedative or anesthetic (≤5 administrations

within 15 min); (2) time to full alertness, defined as the time from surgery completion to the first MOAA/S score of 5 for three consecutive evaluations; (3) time to operating room discharge, defined as the time from surgery completion to a Steward score ≥4; and (4) total dosage of anesthetic agents administered.

Safety assessments

Hemodynamic and respiratory parameters (SBP, DBP, MAP, HR, SpO_2, and RR) were recorded at five time points: before anesthesia (T0), 2 min after induction (T1), at the start of surgery (T2), at the end of surgery (T3), and during awakening (T4). AEs were monitored throughout the study, including hypoxia (SpO_2<95% for ≥30 s), bradycardia (HR <55 beats/min for ≥2 min), and hypotension (SBP<90 mmHg or ≥20% decrease from baseline for ≥2 min). The incidence of intraoperative body movement, injection pain, and postoperative complications—such as dizziness, postoperative nausea, and vomiting (PONV)—was documented. Satisfaction with sedation and anesthesia was assessed separately for patients and surgeons using structured questionnaires.

Statistical analysis

Sample size estimation was based on prior clinical research involving propofol, assuming an anesthesia onset time of 1.2±0.6 min. With a significance level of 0.05, statistical power of 90%, and a non-inferiority margin of 0.2 min, 310 participants were required. Considering a 10% dropout rate, 342 patients were ultimately enrolled.

Statistical analyses were performed using Statistical Package for the Social Sciences Statistics version 25.0 (IBM, USA). Continuous variables were expressed as mean±standard deviation (SD) and compared using the unpaired t-test or Wilcoxon rank-sum test, as appropriate. Repeated-measures Analysis of Variance was applied for hemodynamic parameters. Categorical variables were presented as frequencies or percentages and analyzed using the chi-square test or Fisher's exact test. A $p<0.05$ was considered statistically significant.

RESULTS

Baseline characteristics of enrolled patients

A total of 342 patients were screened for eligibility. One patient in the propofol group was excluded due to a change in anesthesia method during surgery, and two patients in the ciprofol group were excluded owing to unfavorable risk–benefit assessment. Ultimately, 339 patients completed the study between August 1, 2022, and August 31, 2023 (Figure 1). Baseline demographic and clinical characteristics were comparable between the two groups. The mean age was 42.0±10.0 years in the propofol group and 43.4±9.5 years in the ciprofol group. Average height and weight were 157.1±4.6 cm and 56.9±7.6 kg in the propofol group, and 157.5±5.0 cm and 57.6±7.4 kg in the ciprofol group, respectively. Regarding ASA classification, 80.6 and 73.4% of patients were ASA I in the propofol and ciprofol groups, respectively, with the remaining 19.4 and 26.6% classified as ASA II. No statistically significant differences were observed between the groups.

Primary outcomes

The primary endpoint was the onset time of anesthesia. The mean onset time was 37.2±10.4 s in the ciprofol group and 38.7±11.2 s in the propofol group ($p=0.214$), indicating no significant difference between treatments (Table 1).

Secondary outcomes

Both groups achieved a 100% anesthesia success rate. The time to full alertness was slightly longer in the ciprofol group (4.6±1.4 min) than in the propofol group (3.8±1.3 min; $p<0.001$), though the difference was clinically acceptable. The time to exit the operating room did not differ significantly (8.5±2.0 min vs. 8.1±3.1 min; $p=0.112$).

The total anesthetic dose was markedly lower in the ciprofol group (68.5±31.7 mg) than in the propofol group (359.6±164.6 mg; $p<0.001$). Regarding satisfaction, both surgeons (100 vs. 88.8%) and patients (98.2 vs. 71.2%) rated the ciprofol group significantly higher ($p<0.001$).

Intraoperative body movement occurred in only four patients (2.4%) receiving ciprofol compared with 26 patients (15.3%) in the propofol group ($p<0.001$), reflecting improved surgical conditions with ciprofol.

Safety assessments

Pain on injection was reported in 36.5% of propofol patients but in none of the ciprofol group ($p<0.001$). Adverse respiratory and circulatory events were also less frequent with ciprofol. Hypoxia occurred in 11.2% of the propofol group versus 2.4% of the ciprofol group ($p=0.001$), while respiratory depression was seen in 3.5% of propofol patients and in none with ciprofol ($p=0.040$). Hypotension was notably lower in the ciprofol group (10.1%) compared with propofol (41.2%; $p<0.001$). Minor dizziness was reported by 2.4% of propofol patients, but none in the ciprofol group, and no cases of bradycardia, nausea, or vomiting occurred in either group.

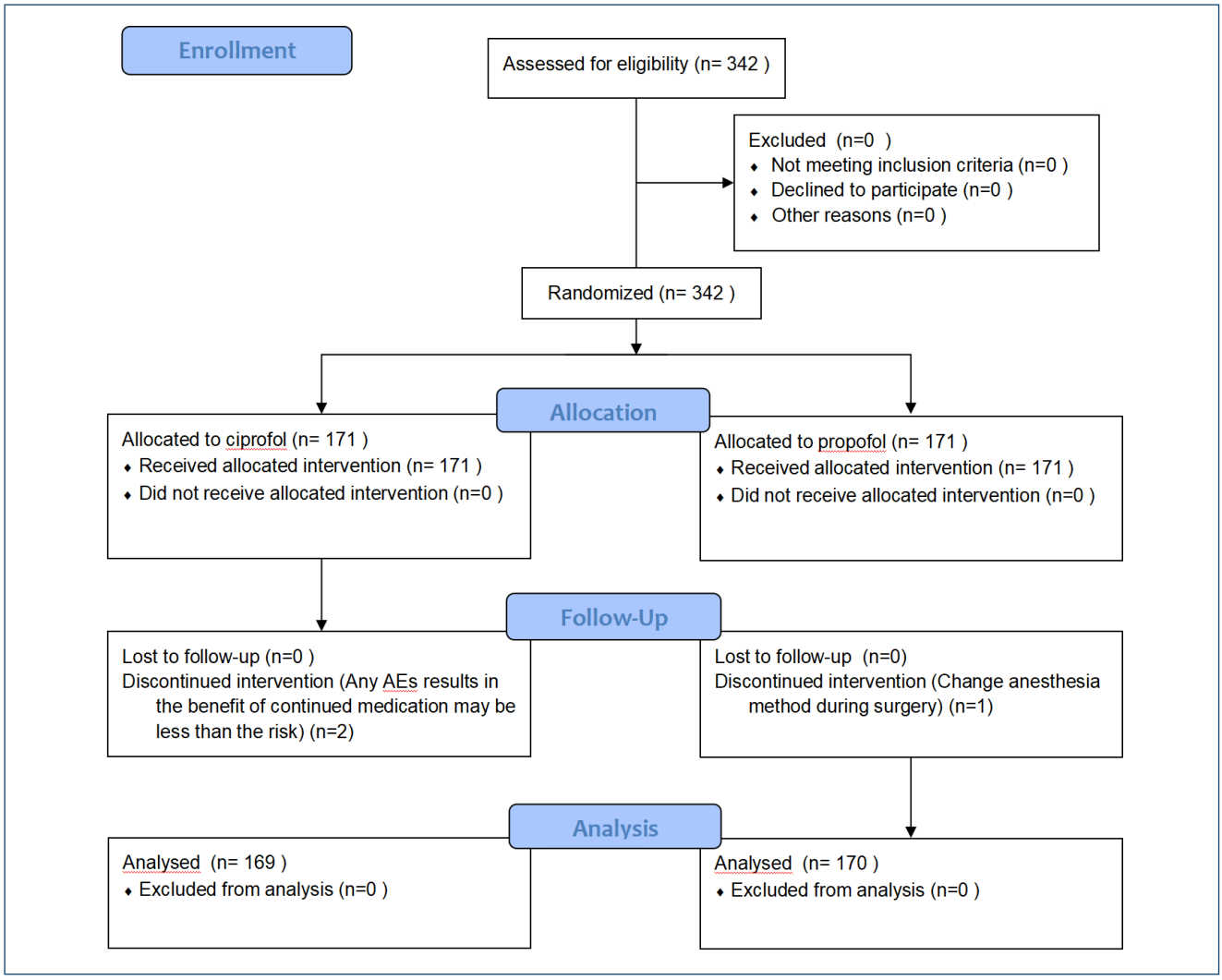


Figure 1. Flow diagram of the CONSORT.

Hemodynamic changes

Both groups exhibited transient reductions in MAP, HR, and SpO_2 after induction (T1–T4) compared with baseline (T0) ($p<0.05$), though all remained within normal clinical limits. The ciprofol group demonstrated smaller fluctuations in MAP, HR, and SpO_2 than the propofol group, indicating superior hemodynamic stability (Figure 2).

DISCUSSION

Patients undergoing minor gynecologic surgery often experience considerable anxiety, highlighting the need for effective and safe sedative agents. This randomized trial demonstrated that ciprofol was non-inferior to propofol in anesthesia onset time during minor gynecologic surgery. Although recovery to full alertness was slightly longer with ciprofol, the delay was clinically acceptable. More importantly, ciprofol offered clear advantages, including less injection pain, fewer AEs, greater hemodynamic stability, and higher patient and surgeon satisfaction. These findings indicate that ciprofol is a promising sedative for short gynecologic procedures.

Propofol remains the most widely used intravenous anesthetic due to its rapid onset, predictable recovery, and high controllability. However, its use is frequently limited by adverse effects such as injection pain, cardiorespiratory depression, and hypotension[6]. Ciprofol, a novel phenol derivative structurally similar to propofol, has demonstrated comparable sedative efficacy but improved safety profiles. Its enhanced affinity for the GABA. A receptor, attributed to the introduction of a cyclopropyl group in its molecular structure, strengthens its anesthetic potency while reducing undesirable effects such as respiratory depression and vascular irritation.

Table 1. Summary of secondary endpoints and adverse events between the two groups.

Variables	Propofol (n=170)	Ciprofol (n=169)	p-value
The onset time of anesthesia (s)	38.7±11.2	37.2±10.4	0.214
Fully alert time (min)	3.8±1.3	4.6±1.4	<0.001
Time out of the operating room (min)	8.1±3.1	8.5±2.0	0.112
Top-up doses required (mg)	359.6±164.6	68.5±31.7	<0.001
Body movement, n (%)	26 (15.3%)	4 (2.4%)	<0.001
Patients satisfaction			
Satisfaction	121 (71.2%)	166 (98.2%)	<0.001
General satisfaction	49 (28.8%)	3 (1.8%)	
Dissatisfaction	0 (0.0%)	0 (0.0%)	
Surgeons satisfaction			
Satisfaction	151 (88.8%)	169 (100%)	<0.001
General satisfaction	19 (11.2%)	0 (0.0%)	
Dissatisfaction	0 (0.0%)	0 (0.0%)	
Adverse events, n (%)			
Hypoxia	19 (11.2%)	4 (2.4%)	0.001
Respiratory depression	6 (3.5%)	0 (0.0%)	0.040
Bradycardia	0 (0.0%)	0 (0.0%)	1.000
Hypotension	70 (41.2%)	17 (10.1%)	<0.001
Pain on injection	62 (36.5%)	0 (0.0%)	<0.001
Dizziness	4 (2.4%)	0 (0.0%)	0.133
Nausea/vomiting	0 (0.0%)	0 (0.0%)	1.000

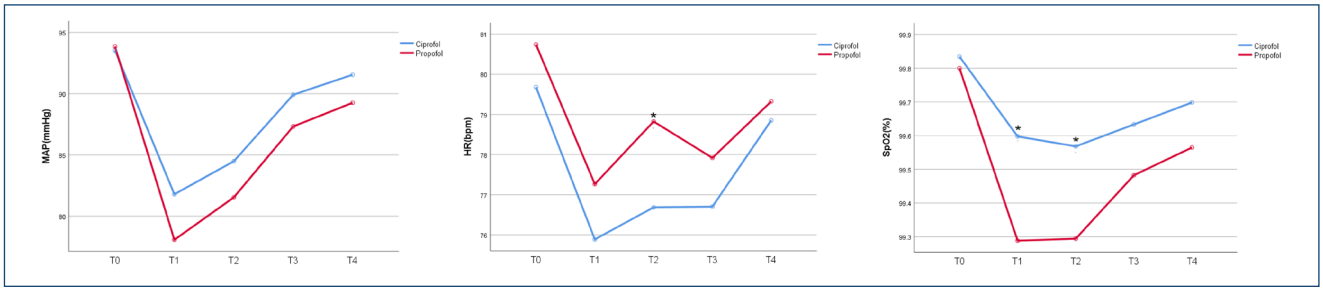


Figure 2. Comparisons of mean arterial pressure, heart rate, and pulse oxygen saturation (SpO_2) between the ciprofol and propofol groups. T0: Pre-anesthesia, T1: 2 min post induction, T2: the beginning of the operation, T3: the end of the operation, T4: awakening.

Consistent with earlier studies, this investigation found no significant differences between ciprofol and propofol in anesthesia onset, anesthesia success rate, or operating room discharge time. The slightly prolonged recovery time with ciprofol aligns with prior findings in patients undergoing gastroscopy and colonoscopy. However, other studies have reported longer onset times for ciprofol, which may be influenced by factors such as patient age, procedure type, or concomitant sedative and analgesic use[7].

In terms of safety, oxygen desaturation—the most frequent AE related to propofol—occurred less often with ciprofol. This observation supports previous evidence that ciprofol causes less respiratory suppression due to its distinct molecular configuration and higher receptor selectivity[8]. This advantage of ciprofol may be attributed to its pharmacological specificity: (1) Ciprofol has a 4–5 fold higher affinity for GABA receptors and can achieve the same sedative effect at lower plasma concentrations when compared to propofol; (2) Ciprofol induces respiratory rhythm disorders through direct medullary inhibition, thereby reducing inhibition of the brainstem respiratory center. Similarly, the incidence of hypotension was significantly lower in the ciprofol group, in agreement with prior multicenter studies[9]. All hypotensive episodes were transient and

self-resolving, indicating favorable hemodynamic tolerance. This may be because ciprofol enhances the selectivity of $GABA_A$ receptor β 3 subunit and more effectively inhibits overactivation of the sympathetic adrenal medulla system.

Higher satisfaction scores among both surgeons and patients further reinforced ciprofol's advantages. Surgeons valued the stable operating conditions and reduced body movements (2.4 vs. 15.3%), while patients appreciated the absence of injection pain (0 vs. 36.5%). These results are consistent with previous clinical reports, confirming that ciprofol's structural modification—specifically the cyclopropane substitution that forms a chiral center—enhanced hydrophobicity, reduces vascular irritation, and eliminates the characteristic injection discomfort observed with propofol. Consequently, ciprofol provides a smoother and more comfortable anesthesia experience for both patients and clinicians.

This study provided high-quality evidence by using a randomized, double-blind, active-controlled design involving a large sample of 339 patients. Standardized sedation protocols and rigorous monitoring ensured consistent data collection. The head-to-head comparison between ciprofol and propofol under identical operative conditions allowed direct evaluation of pharmacodynamic differences. Moreover, multiple clinically meaningful endpoints—including onset time, recovery, hemodynamic parameters, adverse respiratory events, and satisfaction scores—provide a comprehensive safety and efficacy profile.

Several limitations should be acknowledged. First, the study population included only ASA I–II patients aged 18–65 years; thus, findings may not be generalizable to elderly, pediatric, or high-risk patients. Second, this was a single-center trial, which may limit external validity despite the large sample size. Third, the assessment of safety of ciprofol was limited to the in-hospital phase, and future studies incorporating patient-reported outcome measures at longer-term follow-ups are warranted. Finally, although both sedation regimens followed standardized dosing, variability in individual pharmacokinetics may still influence onset and recovery.

The study's findings have direct and practical clinical relevance. Ciprofol demonstrated equivalent anesthetic effectiveness but significantly better safety, particularly regarding reduced hypotension, minimal respiratory suppression, and complete elimination of injection pain. These advantages make ciprofol especially suitable for outpatient gynecologic procedures, where rapid turnover, patient comfort, and hemodynamic stability are paramount. Improved surgeon and patient satisfaction further supports integration of ciprofol as a preferred sedative agent in ambulatory gynecologic anesthesia pathways.

CONCLUSION

Ciprofol provides anesthetic efficacy comparable to propofol while offering markedly superior safety, stability, and patient experience during minor gynecologic surgery. Its reduced incidence of hypotension, respiratory compromise, and injection pain, combined with high satisfaction ratings, supports its use as a preferred sedative in fast-turnover outpatient gynecologic settings. Future multicenter studies across broader patient populations will help refine its clinical role.

ETHICAL STATEMENT

This trial was conducted in accordance with the Declaration of Helsinki and the CONSORT 2010 guidelines. The study was approved by the Clinical Research Ethics Committee of Heyuan People's Hospital (YXLL-2021K35), which was registered at http://www.chictr.org.cn (13/12/2022, ChiCTR2200066673). Written informed consent was obtained from patients undergoing minor gynecologic surgery at Heyuan People's Hospital from 01/08/2022 to 31/08/2023.

AUTHORS' CONTRIBUTIONS

†**WH**, **QC**, and **DP** contributed equally to this work. **WH:** Conceptualization, Data curation, Writing – original draft, Writing – review & editing. **QC:** Conceptualization, Data curation, Writing – review & editing. **DP:** Conceptualization, Data curation, Writing – review & editing. **YL:** Data curation, Writing – review & editing. **WL:** Writing – review & editing. **SP:** Writing – review & editing. **GX:** Conceptualization, Writing – original draft, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Glen JB. Try, try, and try again: personal reflections on the development of propofol. Br J Anaesth. 2019;123(1):3-9. https://doi.org/10.1016/j.bja.2019.02.031

2. Qin L, Ren L, Wan S, Liu G, Luo X, Liu Z, et al. Design, synthesis, and evaluation of novel 2,6-disubstituted phenol derivatives as general anesthetics. J Med Chem. 2017;60(9):3606-17. https://doi.org/10.1021/acs.jmedchem.7b00254

3. Liao J, Li M, Huang C, Yu Y, Chen Y, Gan J, et al. Pharmacodynamics and pharmacokinetics of HSK3486, a novel 2,6-disubstituted phenol derivative as a general anesthetic. Front Pharmacol. 2022;13:830791. https://doi.org/10.3389/fphar.2022.830791

4. Luo Z, Tu H, Zhang X, Wang X, Ouyang W, Wei X, et al. Efficacy and safety of HSK3486 for anesthesia/sedation in patients undergoing fiberoptic bronchoscopy: a multicenter, double-blind, propofol-controlled, randomized, phase 3 study. CNS Drugs. 2022;36(3):301-13. https://doi.org/10.1007/s40263-021-00890-1

5. Wang X, Wang X, Liu J, Zuo YX, Zhu QM, Wei XC, et al. Effects of ciprofol for the induction of general anesthesia in patients scheduled for elective surgery compared to propofol: a phase 3, multicenter, randomized, double-blind, comparative study. Eur Rev Med Pharmacol Sci. 2022;26(5):1607-17. https://doi.org/10.26355/eurrev_202203_28228

6. Ainiwaer D, Jiang W. Efficacy and safety of ciprofol versus propofol for anesthesia induction in adult patients received elective surgeries: a meta-analysis. BMC Anesthesiol. 2024;24(1):93. https://doi.org/10.1186/s12871-024-02479-9

7. Xu J, Yang M, Zeng Y, Zou XH, Ren JH, Xia Z, et al. Efficacy and safety of ciprofol for sedation in outpatient gynecological procedures: a phase III multicenter randomized trial. Front Med (Lausanne). 2024;11:1360508. https://doi.org/10.3389/fmed.2024.1360508

8. Lan H, Shan W, Wu Y, Xu Q, Dong X, Mei P, et al. Efficacy and safety of ciprofol for sedation/anesthesia in patients undergoing hysteroscopy: a randomized, parallel-group, controlled trial. Drug Des Devel Ther. 2023;17:1707-17. https://doi.org/10.2147/DDDT.S414243

9. Zhang J, Liu R, Bi R, Li X, Xu M, Li L, et al. Comparison of ciprofol-alfentanil and propofol-alfentanil sedation during bidirectional endoscopy: a prospective, double-blind, randomised, controlled trial. Dig Liver Dis. 2024;56(4):663-71. https://doi.org/10.1016/j.dld.2023.09.016

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251804

Determining the relationship between breast cancer fear in women and levels of cyberchondria

Neşe Kıskaç[1], Merve Özerdem[2], Refika Karahan[3], Dilan Orhan[2], Muharrem Kıskaç[4]*

SUMMARY

OBJECTIVE: The aim of this study was to examine the relationship between breast cancer fear and the level of cyberchondria in women.

METHODS: This descriptive and cross-sectional study was conducted with 379 women aged 18 years and older. Data were collected using a Personal Information Form, the Cyberchondria Severity Scale-Short Form, and the Breast Cancer Fear Scale. Statistical analyses were performed using IBM Statistical Package for the Social Sciences version 26.0.

RESULTS: The participants' mean total score on the Cyberchondria Severity Scale-Short Form was 33.73±10.72 (moderate level), while the mean total score on the Breast Cancer Fear Scale was 24.37±9.11 (high level). Significant associations were found between the Cyberchondria Severity Scale-Short Form total score and participants' educational status, menopausal status, and practice of breast self-examination, as well as between the Breast Cancer Fear Scale total score and history of clinical breast examination ($p<0.05$). A positive significant correlation was observed between the Cyberchondria Severity Scale-Short Form total score and the Breast Cancer Fear Scale total score ($r=0.464$; $p<0.05$).

CONCLUSION: Nurses should promote health literacy and guide individuals toward reliable sources, employing an interdisciplinary approach to manage patient fears. By increasing awareness of regular screenings and breast self-examinations, nurses play a key role in mitigating anxiety and fear among high-risk women.

KEYWORDS: Breast cancer. Anxiety. Fear.

INTRODUCTION

Cancer remains a global public health priority as the second leading cause of death, with incidences rising due to aging populations and lifestyle factors[1,2]. Among women, breast cancer is the most prevalent malignancy, accounting for 24% of all female cancer cases with 2.3 million annual diagnoses[3]. Beyond physiological effects, a breast cancer diagnosis significantly impairs psychological well-being, often triggering fear, anxiety, and uncertainty regarding recurrence. These fatalistic perceptions influence coping mechanisms and health behaviors, frequently driving individuals to seek information[4]. In the digital era, the internet is a primary source of health information; however, excessive searching can lead to cyberchondria, an escalation of anxiety resulting from online health inquiries[5]. Exposure to inaccurate digital information can undermine treatment trust and provoke unnecessary panic[6-11]. Consequently, nurses carry a substantial responsibility to provide professional support, promote early detection, and guide patients toward reliable sources to mitigate anxiety[4]. This study aims to examine the relationship between breast cancer fear and cyberchondria levels in women. Understanding this interaction is crucial for improving health behaviors and guiding healthcare professionals in developing strategies to manage the psychological impact of the disease, ultimately enhancing the quality of life for those at risk or diagnosed.

METHODS

Study design

This research was designed as a descriptive and cross-sectional study.

Sample

The study's sample size was calculated as 377 based on a 95%CI and a 5% margin of error, with 379 women ultimately participating. Utilizing snowball sampling, participants were recruited via telephone directories. Data collection was conducted digitally through Google Forms sent to the participants' mobile devices.

[1]Istanbul Arel University, Faculty of Health Sciences, Department of Nursing – Istanbul, Turkey.
[2]Istanbul Gelisim University, Faculty of Health Science, Department of Nursing – Istanbul, Turkey.
[3]Acıbadem Fulya Hospital – Istanbul, Turkey.
[4]Biruni University, Faculty of Medicine, Department of Internal Medicine – Istanbul, Turkey.
*Corresponding author: muharremkiskac@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on November 26, 2025. Accepted on January 30, 2026.
Scientific Editor: José Maria Soares Júnior

Data collection procedure

Data were collected online via a self-administered questionnaire between January 1 and April 9, 2025.

Instruments

Study data were obtained using a Personal Information Form, the Cyberchondria Severity Scale-Short Form (CSS-SF), and the Breast Cancer Fear Scale (BCFS). Guidelines on breast cancer list the individual's age, age at first menstruation, menopausal status, oral contraceptive use, body mass index (BMI), age at first birth, family history of breast cancer, hormone replacement therapy, and exercise as risk factors for breast cancer. Therefore, the personal data collection form for women collects information specified as risk factors in the guidelines[12,13].

Cyberchondria Severity Scale-Short Form

Cyberchondria involves heightened anxiety from excessive online health searching. We used the 12-item Short Form (CSS-SF), developed by McElroy and Shevlin and revised by McElroy et al.[14,15]. Validated in Turkish[16,17], it assesses four subscales on a five-point Likert scale (score range: 12–60). Higher scores indicate greater severity. This study found a Cronbach's alpha of 0.89.

Breast Cancer Fear Scale

The 8-item BCFS, developed by Champion et al.[18] and validated in Turkish by Seçginli[19], assesses emotional responses to breast cancer. Scores (8–40) categorize fear as low, moderate, or high. In this study, Cronbach's alpha was 0.93, indicating excellent internal consistency.

Data analysis

Data were analyzed using IBM Statistical Package for the Social Sciences 26.0. To compare groups, t-test/analysis of variance (for normal distribution) and Mann-Whitney U/Kruskal-Wallis (for non-normal distribution) were utilized. Relationships were assessed via Pearson/Spearman correlations, while multiple regression controlled for demographic variables. Statistical significance was set at $p<0.05$ with 95%CI.

Ethical considerations

Ethical approval was granted by Istanbul Gelisim University Ethics Committee (No: 2024-20-10; Date: 19 Dec 2024). All participants provided written informed consent after being briefed on the study's purpose. This research adhered to the Declaration of Helsinki and national ethical standards.

RESULTS

The relationships between participants' sociodemographic characteristics, the total scores of the CSS-SF, and the BCFS are presented in Tables 1 and 2. The mean total CSS-SF score of the participants was 33.73±10.72, indicating a moderate level of cyberchondria. The mean total BCFS score was 24.37±9.11, reflecting a high level of breast cancer fear. Significant associations were found between the CSS-SF total score and participants' educational status, menopausal status, and practice of breast self-examination, as well as between the BCFS total score and history of undergoing clinical breast examination ($p<0.05$) (Tables 1 and 2). CSS-SF scores significantly associated with educational status, menopausal status, and breast self-examination ($p<0.05$). To control for these confounders, multiple regression was performed. Results showed that independent of education ($p=0.665$), menopause ($p=0.735$), and self-examination ($p=0.077$), a moderate positive correlation persisted between total BCFS and CSS-SF scores ($r=0.464$; $p<0.05$).

A statistically significant positive correlation was found between the CSS-SF total score and the BCFS total score ($r=0.464$; $p<0.05$). Furthermore, a significant positive correlation was also found between the BCFS and CSS-SF subdimensions (Table 3).

DISCUSSION

This study aims to determine the relationship between breast cancer fear and cyberchondria in women, and to identify the associated individual characteristics. A review of the literature reveals a limited number of studies examining the link between cancer and cyberchondria; however, no research has been found specifically investigating the relationship between breast cancer fear and cyberchondria among women. Therefore, the findings of this study will be discussed in the context of the limited existing literature.

This study identified a moderate positive correlation between breast cancer fear and cyberchondria ($r=0.464$; $p<0.05$), with the "compulsions" subscale showing the strongest association ($r=0.436$). Findings suggest online research exacerbates fear, aligning with literature. Coşkun et al. noted that while extensive health-related searching increases awareness, it induces confusion and correlates with compulsive information-seeking[20]. Similarly, Topkara Sucu et al. found significantly higher cyberchondria severity in adolescents with Polycystic Ovary Syndrome compared to healthy controls[21]. Öncü et al. also reported a significant correlation between Human Papillomavirus 16/18 positivity and cyberchondria levels[22]. Consistent with these studies, our results indicate that medical conditions increase

cyberchondria severity; although this may enhance awareness, it simultaneously triggers confusion, ultimately escalating anxiety and fear.

In both the current study and the study by Öncü et al.[22], cyberchondria severity was found to be significantly higher among individuals with higher education levels. It is considered that those with higher education may possess greater awareness of risk factors, which potentially accounts for their elevated cyberchondria levels. Conversely, in the study by Coşkun et al., although no significant correlation was found between education level and cyberchondria severity, scores were notably higher among individuals with lower education levels[20]. This discrepancy may be attributed to the fact that lower levels of health literacy or awareness might lead individuals to engage in more extensive online research.

The significantly lower CSS-SF scores in postmenopausal women may be attributed to unique psychosocial stressors of this stage. Citing Zangirolami-Raimundo et al., who highlighted

Table 1. Relationship between participants' sociodemographic characteristics and the total scores of the Cyberchondria Severity Scale-Short Form and the Breast Cancer Fear Scale (n=379).

	n	%	Total CSS-SF score	p	Total BCFS score	p
Age (years)	30.65±11.35					
≥30	150	39.6	33.35±11.29	0.577	24.09±9.44	0.652
<30	228	60.4	33.98±10.37		24.53±8.90	
Marital status						
Single	228	60.2	34.18±10.06	0.315	24.39±9.06	0.976
Married	151	39.8	33.05±11.65		24.36±9.21	
Educational level						
Primary and secondary school	45	11.9	30.13±11.08	**0.006**	21.98±10.22	0.254
High school	84	22.2	32.00±12.02		24.08±9.71	
Undergraduate	222	58.6	34.65±10.25		24.94±8.75	
Postgraduate	28	7.4	37.43±7.25		24.61±7.86	
Age at menarche	13.36±1.87					
≥13	266	70.2	33.50±10.71	0.510	24.19±9.16	0.549
<13	113	29.8	34.29±10.77		24.81±9.00	
Menopausal status						
No	346	91.3	34.12±10.48	**0.023**	24.52±9.08	0.324
Yes	33	8.7	29.70±12.43		22.88±9.35	
Age at menopause	46.79±4.65 (n=33)					
≥46	20	60.6	29.62±12.32	0.904	23.38±9.66	0.678
<46	13	39.4	30.15±12.66		22.00±8.75	
Age at first birth	24.51±4.47 (n=139)					
≥24	78	56.1	33.41±12.23	0.916	24.56±9.78	0.710
<24	61	43.9	33.62±11.15		23.95±9.43	
Use of oral contraceptives						
No	338	89.2	33.78±10.73	0.805	24.61±9.05	0.150
Yes	41	10.8	33.34±10.77		22.44±9.43	
Body mass ındex	24.89±4.49					
≥24	194	51.2	32.99±11.37	0.167	24.07±9.70	0.509
<24	185	48.8	34.51±9.97		24.69±8.46	

Descriptive statistical methods (mean, standard deviation, frequency, and percentage) were used. For group comparisons, Student's t-test and Mann-Whitney U test were applied as appropriate. CSS-SF: Cyberchondria Severity Scale-Short Form; BCFS: Breast Cancer Fear Scale. Bold text indicates a statistically significant difference between the variables ($p<0.05$).

depression during the climacteric period, it is suggested that psychological shifts across different life stages influence health information-seeking behaviors[23].

Participants practicing breast self-examination exhibited significantly higher CSS-SF scores. Research indicates that emotional factors and fears can modify individuals' clinical outlooks and health-related behaviors. In this study, it is suggested that the fear experienced by women acted as a driver, altering their health behaviors and increasing the frequency of Breast Self-Examination practice[24,25].

Although not always reaching statistical significance across all parameters, cyberchondria scores were found to be higher among participants under the age of 30, those who were premenopausal ($p<0.05$), and those who performed breast self-examination ($p<0.05$). Additionally, a low-level significant correlation was identified between age and the "excessiveness" subdimension of the cyberchondria scale. These findings suggest that individuals under the age of 30 engage in more intensive health-related online research. Consequently, increased awareness regarding breast cancer may have prompted a higher frequency of breast self-examination behaviors. Multiple regression analysis conducted in this study revealed that, independent of demographic variables, there is a moderate, positive, and significant relationship between cyberchondria scores and breast cancer fear. In alignment with

Table 2. Relationship between participants' sociodemographic characteristics and the total scores of the Cyberchondria Severity Scale-Short Form and the Breast Cancer Fear Scale (n=379) continue.

	n	%	Total CSS-SF score	P	Total BCFS score	p
Hormone replacement therapy (HRT) use						
No	360	95.0	33.80±10.66	0.615	24.32±9.12	0.626
Yes	19	5.0	32.53±12.10		25.37±9.02	
Regular exercise status						
No	265	69.9	34.02±10.84	0.430	24.44±9.17	0.838
Yes	114	30.1	33.07±10.46		24.23±9.00	
History of breast-related health problems						
No	348	91.8	33.76±10.71	0.865	24.45±9.11	0.613
Yes	31	8.2	33.42±10.98		23.58±9.17	
Family history of breast cancer						
No	306	80.7	33.53±10.74	0.456	23.92±9.01	0.470
Yes	73	19.3	34.58±10.69		26.27±9.31	
Practice of breast self-examination (BSE)						
No	144	38.0	31.86±10.28	**0.008**	24.33±8.74	0.936
Yes	235	62.0	34.88±10.85		24.40±9.34	
History of undergoing mammography						
No	306	80.7	33.74±10.72	0.985	24.37±9.21	0.981
Yes	73	19.3	33.71±10.79		24.40±8.71	
History of receiving education about breast cancer						
No	263	69.4	33.46±10.63	0.456	24.56±9.01	0.562
Yes	116	30.6	34.35±10.94		23.97±9.35	
Total CSS-SF score	33.73±10.72 (min:12–max:60)					
Excessiveness	10.24±3.59 (min:3–max:15)					
Distress score	8.28±3.37 (min:3–max:15)					
Compulsions	8.95±3.18 (min:3–max:15)					
Difficulty score	6.26±3.07 (min:3–max:15)					
Total BCFS score	24.37±9.11 (min:8–max:40)					

Descriptive statistical methods (mean, standard deviation, frequency, and percentage) were used. For group comparisons, Student's t-test and Mann-Whitney U test were applied as appropriate. CSS-SF: Cyberchondria Severity Scale-Short Form; BCFS: Breast Cancer Fear Scale; HRT: Hormone replacement therapy; BSE: breast self-examination. Bold text indicates a statistically significant difference between the variables ($p<0.05$).

the existing literature, this study demonstrates that the severity of cyberchondria exacerbates the fear of breast cancer.

Limitations

This study has several limitations. The use of snowball sampling and an online survey may limit the generalizability of the findings to the broader population of women, potentially over-representing those with higher digital literacy or health anxiety. The cross-sectional design cannot establish causality between breast cancer fear and cyberchondria. Future longitudinal studies with more representative sampling methods are needed to confirm these associations and explore causal pathways.

CONCLUSION

This study demonstrates a significant relationship between breast cancer fear and cyberchondria levels among women. Particularly in individuals with high education and awareness levels, existing medical conditions or high-risk factors prompt more intensive research; this, in turn, may lead to confusion and further escalate anxiety and stress. These findings underscore the necessity of strengthening digital health literacy within nursing practices. Nurses should educate individuals on health literacy and direct them toward reliable information sources. Furthermore, an interdisciplinary approach involving other healthcare professionals is essential to help individuals manage their fears effectively. By raising awareness about regular screenings and breast self-examination among high-risk women, nurses can play a pivotal role in supporting patients in managing their anxiety and fear.

ETHICS COMMITTEE APPROVAL

Ethical approval for the study was obtained from the Ethics Committee of Istanbul Gelisim University (Decision No: 2024-20-10; Date: 19 December 2024).

Table 3. Relationship between participants' age, body mass index, age at first birth, age at menopause, age at menarche, total Cyberchondria Severity Scale-Short Form score and subdimensions, and total Breast Cancer Fear Scale score (n=379).

		Excessiveness	Compulsions	Distress	Reassurance	CSS-SF	BCFS
Age	r	-0.120	0.030	0.017	0.012	-0.022	-0.024
	p	**0.020**	0.564	0.743	0.816	0.665	0.645
Age at menarche	r	-0.070	-0.092	-0.057	0.014	-0.066	0.013
	p	0.171	0.074	0.270	0.792	0.203	0.804
Age at menopause	r	-0.080	0.099	-0.018	-0.053	-0.019	-0.056
	p	0.653	0.576	0.920	0.765	0.916	0.753
Age at first birth	r	-0.035	-0.038	-0.049	-0.136	-0.076	-0.054
	p	0.680	0.661	0.565	0.110	0.374	0.527
BMI	r	-0.180	0.014	-0.071	0.009	-0.074	-0.004
	p	**0.000**	0.780	0.170	0.856	0.149	0.935
Excessiveness	r	1	0.533	0.650	0.365	0.801	0.286
	p		**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Compulsions	r	0.533	1	0.630	0.560	0.841	0.436
	p	**0.000**		**0.000**	**0.000**	**0.000**	**0.000**
Distress	r	0.650	0.630	1	0.516	0.861	0.382
	p	**0.000**	**0.000**		**0.000**	**0.000**	**0.000**
Reassurance	r	0.365	0.560	0.516	1	0.738	0.409
	p	**0.000**	**0.000**	**0.000**		**0.000**	**0.000**
CSS-SF	r	0.801	0.841	0.861	0.738	1	0.464
	p	**0.000**	**0.000**	**0.000**	**0.000**		**0.000**
BCFS	r	0.286	0.436	0.382	0.409	0.464	1
	p	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	

Pearson's and Spearman's correlation analyses. CSS-SF: Cyberchondria Severity Scale-Short Form; BCFS: Breast Cancer Fear Scale; BMI: body mass index. Bold text indicates a statistically significant difference between the variables ($p<0.05$).

AUTHORS' CONTRIBUTIONS

NK: Conceptualization, Data curation, Formal Analysis, Writing – original draft. **MÖ:** Data curation. **RK:** Data curation. **DO:** Data curation. **MK:** Formal Analysis, Writing – original draft.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. World Health Organization. Cancer fact sheet. 2025. [cited on 2025 Jan 21]. Available from: https://www.who.int/news-room/fact-sheets/detail/cancer
2. American Cancer Society. Global cancer facts & figures. 5th ed. 2022; p. 1-48.
3. Nacar G, Firat ZN, Taşhan ST. The relationship between breast cancer anxiety levels and orthorexia (obsession with healthy eating). J Soc Analyt Health. 2022;2(3):227-32.
4. Ersin F, Dedeoğlu GK. The effect of nurses' perceptions of fear and fatalism regarding breast cancer on their breast cancer early detection behaviors. Gevher Nesibe J Med Health Sci. 2020;5(9):146-53.
5. Altındiş S, İnci BM, Aslan GF, Altındiş M. Examination of levels of cyberchondria among university employees and related factors. Sakarya Tıp Dergisi. 2018;8(2):359-70. https://doi.org/10.31832/smj.423652
6. Özkan O, Sungur C, Özer Ö. Investigation of cyberchondria level and digital literacy on women in Turkey. J Hum Behav Soc Environ. 2022;32(6):768-80. https://doi.org/10.1080/10911359.2021.1962776
7. Petersen A, Tanner C, Munsie M. Citizens' use of digital media to connect with health care: socio-ethical and regulatory implications. Health (London). 2019;23(4):367-84. https://doi.org/10.1177/1363459319847505
8. Nes AAG, Steindal SA, Larsen MH, Heer HC, Lærum-Onsager E, Gjevjon ER. Technological literacy in nursing education: a scoping review. J Prof Nurs. 2021;37(2):320-34. https://doi.org/10.1016/j.profnurs.2021.01.008
9. Terry J, Davies A, Williams C, Tait S, Condon L. Improving the digital literacy competence of nursing and midwifery students: a qualitative study of the experiences of NICE student champions. Nurse Educ Pract. 2019;34:192-8. https://doi.org/10.1016/j.nepr.2018.11.016
10. Harerimana A, Mtshali NG. Using exploratory and confirmatory factor analysis to understand the role of technology in nursing education. Nurse Educ Today. 2020;92:104490. https://doi.org/10.1016/j.nedt.2020.104490
11. Harerimana A, Duma SE, Mtshali NG. First-year nursing students' digital literacy: a cross-sectional study. J Nurs Educ Pract. 2023;13(1):31-7.
12. National Comprehensive Cancer Network Guidelines. Breast cancer screening and diagnosis. Version 2.2025. 2025. [cited on 2025 Dec 5]. Available from: https://www.nccn.org/guidelines/guidelines-detail?category=2&id=1421
13. Ministry of Health General Directorate of Health Services. Breast cancer prevention, screening, diagnosis, treatment, and follow-up clinical guidelines. 2024;1170:7-8.
14. McElroy E, Shevlin M. The development and initial validation of the cyberchondria severity scale (CSS). J Anxiety Disord. 2014;28(2):259-65. https://doi.org/10.1016/j.janxdis.2013.12.007
15. McElroy E, Kearney M, Touhey J, Evans J, Cooke Y, Shevlin M. The CSS-12: development and validation of a short-form version of the cyberchondria severity scale. Cyberpsychol Behav Soc Netw. 2019;22(5):330-5. https://doi.org/10.1089/cyber.2018.0624
16. Tuğtekin U, Tuğtekin EB. Siberkondri Ciddiyet Ölçeği'nin kısa formunun Türkçeye uyarlanması ve öğretmen adaylarının aşırı çevrim içi bilgi arama davranışları. Anemon Muş Alparslan Univ Sos Bilim Derg. 2021;9(6):1747-62. https://doi.org/10.18506/anemon.963253
17. Tarcan GY, Karahan A, Sebik NB. Kısa Form Siberkondri (CSS-12) Ciddiyet Ölçeği'nin geçerlik ve güvenirliği: sağlık bilişimi alanına özel bir uygulama. Hacettepe Sağlık İdaresi Derg. 2023;26(1):207-18.
18. Champion VL, Skinner CS, Menon U, Rawl S, Giesler RB, Monahan P, et al. A breast cancer fear scale: psychometric development. J Health Psychol. 2004;9(6):753-62. https://doi.org/10.1177/1359105304045383
19. Secginli S. Mammography self-efficacy scale and breast cancer fear scale: psychometric testing of the Turkish versions. Cancer Nurs. 2012;35(5):365-73. https://doi.org/10.1097/NCC.0b013e3182331a9a
20. Coşkun M, Ünlü Suvari G, Özdemir İN. Role of cyberchondria severity in the relationship between HPV awareness and vaccination status among adult women. BMC Public Health. 2025;26(1):357. https://doi.org/10.1186/s12889-025-25938-5
21. Topkara Sucu S, Kolomuç Gayretli T, Küçükkayıkçı AS, Kurt D, Ergörün SK, Eryılmaz T, et al. Cyberchondria levels in adolescent patients with polycystic ovary syndrome in the digital age. J Pediatr Adolesc Gynecol. 2024;37(6):569-73. https://doi.org/10.1016/j.jpag.2024.08.006
22. Öncü HN, Ege G, Sucu ST, Öztürk N, Hanedan C, Korkmaz V. The impact of human papillomavirus 16/18 positivity on cyberchondria levels in women undergoing colposcopy. Rev Assoc Med Bras (1992). 2025;71(5):e20241692. https://doi.org/10.1590/1806-9282.20241692
23. Zangirolami-Raimundo J, Sorpreso ICE, Rebouças CMP, Bezerra PCL, Costa LMPRD, Baracat EC, et al. Depression in women in climacteric period: a brief review. Rev Assoc Med Bras (1992). 2023;69(7):e20230385. https://doi.org/10.1590/1806-9282.20230385
24. Karnatovskaia LV, Johnson MM, Varga K, Highfield JA, Wolfrom BD, Philbrick KL, et al. Stress and fear: clinical ımplications for providers and patients (in the time of COVID-19 and beyond). Mayo Clin Proc. 2020;95(11):2487-98. https://doi.org/10.1016/j.mayocp.2020.08.028
25. Galhardo APM, Andrade PR, Andrade LP, Cury MAA, Mukai MK, Baracat EC, et al. Cross-sectional study of self-reported pain related to temporomandibular disorders and emotional state of medical school faculty and students: post-COVID-19 pandemic. PLoS One. 2024;19(8):e0308988. https://doi.org/10.1371/journal.pone.0308988

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251808

Bullying among medical residents in a large public hospital: a cross-sectional study

Luiz Antônio de Oliveira[1], Aleida Nazareth Soares[1], Renata Rocha Barreto Giaxa[2], Rosa Malena Delbone[1], Alexandre Sampaio Moura[1,2]*

SUMMARY

OBJECTIVE: Bullying is a prevalent issue in medical residency worldwide, yet data from Latin America remain limited. The aim of this study was to assess the prevalence of bullying and its associated factors among medical residents in a public hospital in Brazil.

METHODS: A cross-sectional study was conducted with 82 medical residents. Bullying was assessed using the Negative Acts Questionnaire-Revised.

RESULTS: There was a prevalence of bullying ranging from 45.1% (Leymann's criteria) to 51.2% (Einarsen, Hoel, and Notelaers criteria). The overall median Negative Acts Questionnaire-Revised score was 33 (interquartile range: 27–42), with no significant difference between first-year residents and later-year residents (28.0 vs. 34.5; p=0.06). However, first-year residents reported lower median scores for some bullying domains, including "Work-Related Bullying" (p=0.04), "Personal and Professional Disqualification" (p=0.03), and "Physical Intimidation" (p=0.02). Only 7% of residents self-reported feeling bullied. No associations were found between bullying and sex or specialty.

CONCLUSION: These findings indicate a high prevalence of bullying in the medical residency programs studied, particularly among more advanced trainees. The large gap between assessed and self-perceived bullying suggests a lack of awareness about what constitutes harassment.

KEYWORDS: Bullying. Risk factors. Medical education. Medical residency. Occupational health.

INTRODUCTION

Bullying refers to repeated and long-lasting behaviors that may cause psychological harm, often manifested through verbal aggression, exclusion from team activities, obstruction of tasks, or disqualification of professional identity[1].

The consequences of bullying can extend beyond personal suffering, leading to financial strain, reduced job performance, organizational dysfunction, and diminished productivity[1].

Bullying in healthcare presents specific characteristics, such as frequent underreporting and the perception of its occurrence as a natural part of training, which perpetuate a cycle of aggression and silence[2]. International studies indicate that bullying is highly prevalent in residency training, varying from 37 to 73%, and women and more advanced trainees tend to be more frequently exposed[2-5].

In Brazil, there are few studies on the prevalence and factors associated with bullying during medical training. A study conducted at a university hospital in São Paulo found that approximately 90% of medical residents experienced some form of bullying during their first year of training, with negative impacts on the educational environment and their mental health. The most common types of bullying identified in this study were social isolation and humiliation[6].

We sought to estimate the prevalence of bullying among medical residents at a large public hospital in Brazil and analyze the factors contributing to its occurrence.

METHODS

Study design and setting

This observational cross-sectional study was conducted between March and May 2024, at Santa Casa BH (SCBH). It was approved by the Research Ethics Committee (CAAE: 76491423.6.0000.5138) and conducted in accordance with the Declaration of Helsinki.

Participants and recruitment

The study included a convenience sample of medical residents enrolled in various residency programs at SCBH. Residents who

[1]Faculdade Santa Casa de Belo Horizonte – Belo Horizonte (MG), Brazil.
[2]Faculdade de Ciências Médicas da Santa Casa de São Paulo – São Paulo (SP), Brazil.
*Corresponding author: alexandremoura@faculdadesantacasabh.edu.br
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on November 02, 2025. Accepted on February 18, 2026.
Scientific Editor: Roseli Nomura

were on vacation, on leave, or completing an external rotation at the time of data collection were excluded.

The sample size was calculated using GPower software (version 3.1.9.7). Considering a bullying prevalence of 63% among residents[2], with a 10% margin of error, the required number of participants was estimated at 76.

All 497 medical residents from the 35 residency programs of the institution were eligible to participate, of whom 82 (16.5%) were included. All participants signed informed consent.

Instruments and procedures

Data were collected using an online questionnaire. Bullying was assessed using the Brazilian version of the Negative Acts Questionnaire-Revised (NAQ-R), which comprises 22 items[7]. To assess bullying, three criteria were applied. The first was Leymann's criterion, which defines bullying as exposure to negative behaviors at least once a week for 6 months or more[8]. This definition emphasizes persistence and repetition as central elements for classifying workplace bullying.

The second criterion, proposed by Einarsen, Hoel, and Notelaers, distinguishes between types and intensities of negative behaviors and classifies individuals as "not bullied," "occasionally exposed to bullying," or "victims of bullying"[9].

The third criterion was based on the self-labeling method proposed by Mikkelsen et al. and is based on a dichotomic response (yes/no) to an item that assesses the respondent's self-perception of being a victim of bullying[10].

To analyze the bullying pattern, we opted to use a validated structure with four factors: (1) work-related bullying; (2) personal bullying; (3) personal and professional disqualification; and (4) physical intimidation[9,11]. The prevalence of bullying for each of these factors was estimated using Leymann's criterion, which involves exposure to at least one item of each factor at least once a week for 6 months or more[8].

Statistical analysis

Descriptive analyses were presented as absolute frequencies and percentages for categorical variables and as medians with interquartile ranges (IQRs) for continuous variables. Group comparisons for continuous measures were conducted with the nonparametric Mann-Whitney U test, while associations between categorical variables were examined using Fisher's exact test. Relationships among bullying-related factors were explored through Spearman's rank-order correlation. Statistical significance was defined as $p<0.05$. Analyses were conducted using SPSS software, version 25.0 (IBM Corp., Armonk, NY, USA).

RESULTS

A total of 82 medical residents participated in the study. There was a female predominance among participants (67.1%), with a mean age of 29.6 years (standard deviation [SD]=4.4). Most participants were enrolled in the General Surgery program (25.6%), followed by Pediatrics (15.8%), Internal Medicine (6.1%), and Anesthesiology (4.8%). A total of 41 residents (50.0%) were enrolled in clinical specialties, 32 (39.0%) in surgical specialties, and 8 (9.7%) in mixed specialties. Regarding year of training, 22 (26.8%) were in the first year (R1), 22 (26.8%) in the second year (R2), and 38 (46.3%) were senior residents (R3 or R4).

The distribution of responses in each item of the NAQ-R is shown in Figure 1. The overall prevalence of bullying in the sample varied depending on the criterion used. A total of 37 residents (45.1%) met Leymann's criteria for bullying, and 42 (51.2%) met the Einarsen, Hoel, and Notelaers criteria (Table 1). Using the self-labeling criterion, only six participants (7.3%) considered themselves victims of bullying; of these, two (2.44%) reported that the bullying occurred "several times per week," and four (4.88%) reported that it occurred daily.

When stratifying the analysis by bullying factor, work-related bullying was the most prevalent (37.8%), followed by personal and professional disqualification (29.3%). A strong positive correlation was observed between these two factors (see Supplementary Table 1).

No significant differences were found in the prevalence of bullying (as defined by Leymann) between men and women (48.1 vs. 43.6%; p=0.700). No association was found between the prevalence of bullying and the year of training (p>0.999) or the specialty type (p=0.896).

The overall median bullying scores for R1s did not differ significantly from those observed for more advanced residents (28.0 [IQR: 24.0–36.0] vs. 34.5 [IQR: 28.0–43.0]; p=0.058). In the analysis stratified by bullying factor, the median "personal bullying" scores for R1s were lower than those of more experienced residents (8.0 [IQR: 7.0–12.0] vs. 10.0 [IQR: 8.0–13.0]; p=0.040). The median scores were also lower for R1s in the "personal and professional disqualification" and "physical intimidation" factors. For work-related bullying, no difference in median scores between R1 and more advanced residents was observed (12.5 [9.0–14.0] vs. 13.0 [11.0–16.0]; p=0.196) (Table 2).

No statistically significant differences in overall bullying scores were found by sex or by specialty type (clinical vs. surgical).

DISCUSSION

The overall prevalence of bullying found in our study was high, but consistent with the estimate of 64.1% reported in a meta-analysis on bullying in medical residency worldwide[12].

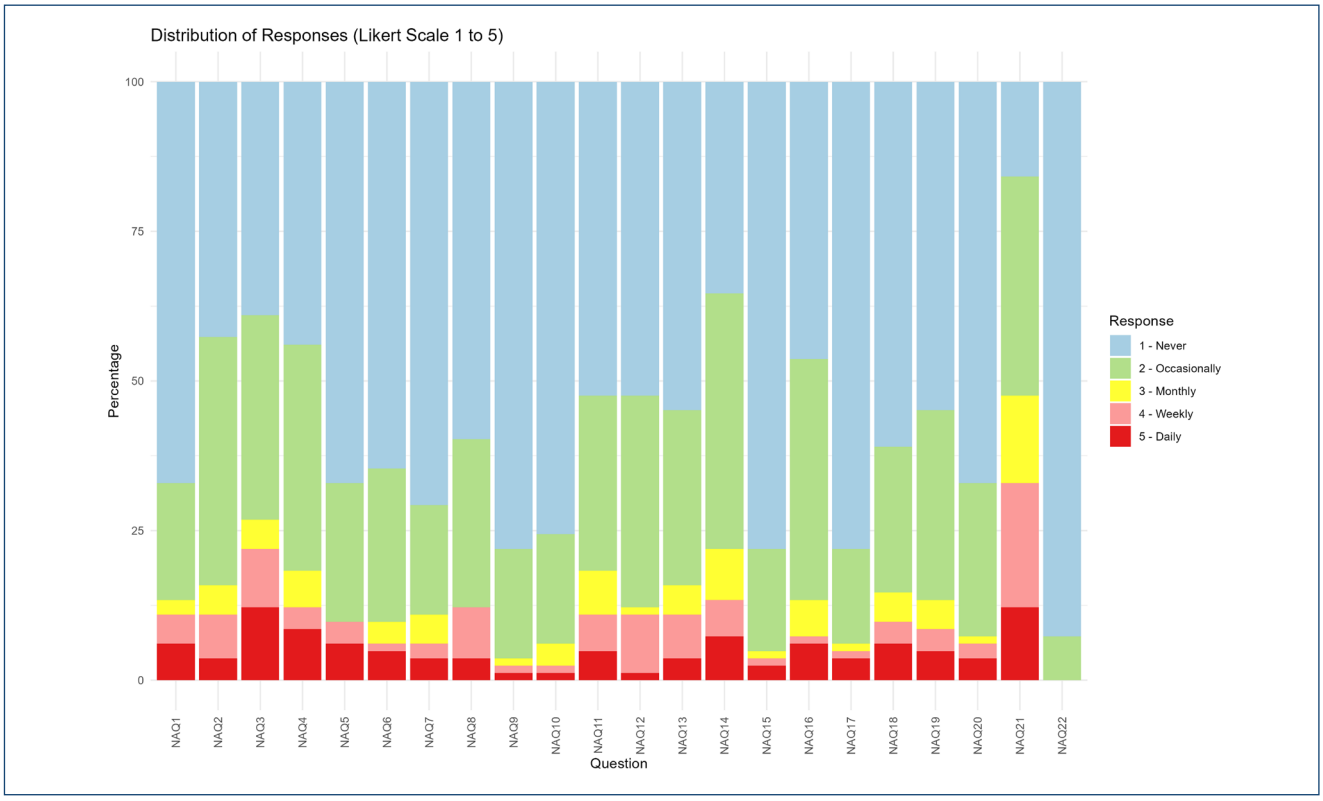


Figure 1. Distribution of responses for each item of the Negative Acts Questionnaire-Revised. Responses are presented as percentages on a 5-point Likert scale.

Table 1. Prevalence of bullying among medical residents according to the criteria proposed by Leymann and Einarsen et al[8,9].

Criteria	Category	n (%)
Leymann[8]	Overall bullying	37 (45.1%)
	Work-related bullying	31 (37.8%)
	Personal bullying	15 (18.3%)
	Personal and professional disqualification	24 (29.3%)
	Physical intimidation	5 (6.1%)
Einarsen et al.[9]	Not bullied	40 (48.8%)
	Occasionally bullied	25 (30.5%)
	Victims of bullying	17 (20.7%)

Our findings were also consistent with prevalence estimates from other studies conducted in Brazil. A study involving medical and multiprofessional residents at a teaching hospital found a prevalence of bullying among physicians of 46.8%[13]. Similarly, a study by the Brazilian College of Surgeons reported a perceived prevalence of bullying of 49.1% during surgical training[14]. Another Brazilian study evaluating general surgery residents found that more than 50% of residents reported experiencing some form of bullying[15].

These findings underscore the necessity for practical measures to mitigate bullying in academic settings, including comprehensive training for the multiprofessional team, the implementation of multiple reporting channels, and fostering an organizational culture of zero tolerance for bullying[16]. Residency program supervisors can help reduce bullying by fostering continuous professional development of clinical staff and establishing clear institutional policies against bullying[17].

One interesting finding of this study was the correlation between different factors of bullying, including work-related bullying, personal and professional disqualification, and physical intimidation. This suggests that bullying behaviors may co-occur, reinforcing the need for an integrated approach to address this issue[1].

Another relevant finding of this study was the marked discrepancy between bullying prevalence estimated using behavior-based criteria and the low proportion of residents who self-labeled as victims of bullying. Such low recognition may be influenced by institutional culture, in which hostile behaviors are normalized or perceived as inherent to medical training, as well as by fear of retaliation in hierarchical environments[18]. Additionally, this discrepancy may reflect a lack of insight into the abusive nature of certain interactions, particularly when bullying occurs in subtle or indirect forms.

Table 2. Bullying scores stratified by factor and residency training year.

Factor	Residency training year	Median	25th percentile	75th percentile	p-value
Overall bullying	R1	28.0	24.0	36.0	0.058
	R2, R3+R4	34.5	28.0	43.0	
Work-related bullying	R1	12.5	9.0	14.0	0.196
	R2, R3+R4	13.0	11.0	16.0	
Personal bullying	R1	8.0	7.0	12.0	**0.040**
	R2, R3+R4	10.0	8.0	13.0	
Personal and professional disqualification	R1	5.5	4.0	8.0	**0.028**
	R2, R3+R4	7.0	5.0	10.0	
Physical intimidation	R1	3.0	3.0	3.0	**0.021**
	R2, R3+R4	4.0	3.0	4.0	

Note: Values are medians and interquartile ranges (25th–75th percentile). The Mann-Whitney U test was used for comparisons. Bold p-values indicate statistical significance ($p<0.05$).

Recent literature in psychology has increasingly focused on the concept of microaggressions[19]. In medical training, microaggressions can reinforce power imbalances and contribute to a culture where certain abusive behaviors are minimized, dismissed, or perceived as part of the learning process. Previous studies have emphasized how microaggressions can erode psychological safety and increase emotional distress among residents, particularly when combined with institutional silence or fear of retaliation[20]. Educational interventions aimed at improving recognition and awareness of workplace harassment have been proposed as strategies to address this gap[21]. Preventing bullying in medical residency may benefit the mental health of medical trainees and could contribute to patient safety and the quality of care[22].

Our results show a similar distribution of bullying occurrences between male and female residents. This finding is consistent with a Brazilian study that also found no significant association between bullying and sex in residency programs[13]. This may reflect a broader trend within medical training environments rather than being necessarily attributed to specific features of organizational culture. Notably, women have increasingly asserted their roles in society and may be less intimidated by traditionally male-dominated environments than they were in the past[23].

We found significant differences between novice and more experienced residents in personal bullying, personal and professional disqualification, and physical intimidation, with first-year residents reporting lower bullying scores. This finding may reflect either reduced exposure to bullying behaviors or lower recognition and reporting of such behaviors among R1s. In addition, they are often perceived as less threatening and more compliant, potentially resulting in fewer confrontational interactions with supervisors or senior colleagues. Limited experience may also contribute to difficulties in recognizing and labeling certain behaviors as bullying, and R1s may interpret hostile or demeaning interactions as inherent to the learning process. In contrast, more senior residents, who assume greater clinical responsibility and leadership roles, may be more frequently subjected to criticism, intimidation, or professional disqualification. These findings are consistent with those of a meta-analysis that identified a positive association between residency seniority and the prevalence of bullying[12].

In our study, we did not find an association between medical specialty and bullying, in line with the findings from a previous study[2]. The authors emphasize that although some surgical specialties are traditionally perceived as having a more hostile environment, the data do not show differences across medical specialties.

A potential limitation of this study is the measurement of bullying. However, the NAQ-R questionnaire has been widely used in research on this topic in medical education. Additionally, the sensitive nature of the topic may have influenced the reported results. A systematic review revealed that residents underreport incidents of workplace or sexual harassment due to fear of retaliation[24]. To minimize the impact of this type of limitation, all data collection and responses were anonymous. Finally, because we used convenience sampling, selection bias might have distorted estimates of bullying prevalence, and the study estimates should be interpreted cautiously. However, Fevre et al. argue that while some people who feel more distressed about workplace relations might be more likely to respond to self-administered questionnaires on the topic, this bias would be offset by a countervailing general tendency to underreport negative workplace behaviors[25].

CONCLUSION

The prevalence of bullying among medical residents was high, mainly because of work-related bullying. First-year residents reported lower levels of bullying across several domains than more advanced trainees. No association was found between bullying and other factors such as sex, age, or specialty type. There was a large gap between assessed and self-perceived bullying, suggesting a lack of awareness about what constitutes harassment.

These findings underscore the responsibility of training institutions to address bullying in medical residency programs actively. Leadership engagement, the implementation of confidential and accessible reporting mechanisms, and the adoption of clear zero-tolerance policies might promote safe learning environments and ensure early recognition and effective management of bullying behaviors.

AUTHORS' CONTRIBUTIONS

LAO: Conceptualization, Data curation, Formal Analysis, Writing – original draft, Writing – review & editing. **ANS:** Formal Analysis, Writing – original draft, Writing – review & editing. **RRBG:** Formal Analysis, Writing – original draft, Writing – review & editing. **RMD:** Formal Analysis, Writing – original draft, Writing – review & editing. **ASM:** Conceptualization, Data curation, Formal Analysis, Writing – original draft, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Nielsen MB, Einarsen S. Outcomes of exposure to workplace bullying: a meta-analytic review. Work Stress. 2012;26(4):309-32. https://doi.org/10.1080/02678373.2012.734709
2. Fnais N, Soobiah C, Chen MH, Lillie E, Perrier L, Tashkhandi M, et al. Harassment and discrimination in medical training: a systematic review and meta-analysis. Acad Med. 2014;89(5):817-27. https://doi.org/10.1097/ACM.0000000000000200
3. Leisy HB, Ahmad M. Altering workplace attitudes for resident education (A.W.A.R.E.): discovering solutions for medical resident bullying through literature review. BMC Med Educ. 2016;16:127. https://doi.org/10.1186/s12909-016-0639-8
4. Álvarez Villalobos NA, León Gutiérrez H, Ruiz Hernandez FG, Elizondo Omaña GG, Vaquera Alfaro HA, Carranza Guzmán FJ. Prevalence and associated factors of bullying in medical residents: a systematic review and meta-analysis. J Occup Health. 2023;65(1):e12418. https://doi.org/10.1002/1348-9585.12418
5. Ayyala MS, Rios R, Wright SM. Perceived bullying among internal medicine residents. JAMA. 2019;322(6):576-8. https://doi.org/10.1001/jama.2019.8616
6. Tokeshi AB, Santos RA, Nogueira-Martins LA, Nunes MDPT, Fidalgo TM. Moral harassment and mental health in medical residents: a longitudinal study. Braz J Psychiatry. 2025;47:e20243579. https://doi.org/10.47626/1516-4446-2024-3579
7. Christ HD, Gauer GJC. Estudo de adaptação e fidedignidade do Questionário de Atos Negativos: Revisado (QAN-R) para o português do Brasil [dissertação]. Porto Alegre (RS): Pontifícia Universidade Católica do Rio Grande do Sul; 2011.
8. Leymann H. The content and development of mobbing at work. Eur J Work Organ Psychol. 1996;5(2):165-84. https://doi.org/10.1080/13594329608414853
9. Einarsen S, Hoel H, Notelaers G. Measuring exposure to bullying and harassment at work: validity, factor structure and psychometric properties of the Negative Acts Questionnaire-Revised. Work Stress. 2009;23(1):24-44. https://doi.org/10.1080/02678370902815673
10. Mikkelsen EG, Einarsen S. Bullying in Danish work-life: prevalence and health correlates. Eur J Work Organ Psychol. 2001;10(4):393-413. https://doi.org/10.1080/13594320143000816
11. Sousa LS, Oliveira RM, Santiago JCDS, Bandeira ÉDS, Brito YCF, Alves HFA, et al. Predictors of moral harassment in nursing work in critical care units. Rev Bras Enferm. 2021;74(3):e20200442. https://doi.org/10.1590/0034-7167-2020-0442
12. Bahji A, Altomare J. Prevalence of intimidation, harassment, and discrimination among resident physicians: a systematic review and meta-analysis. Can Med Educ J. 2020;11(1):e97-123. https://doi.org/10.36834/cmej.57019
13. Marques RC, Martins Filho ED, Paula GS, Santos RR. Assédio moral nas residências médica e não médica de um hospital de ensino. Rev Bras Educ Med. 2012;36(3):401-6. https://doi.org/10.1590/S0100-55022012000500015
14. Santos EG, Roque L, Maya MC, Moreira RC, Lima FL, Correia MITD. perception of harassment among female surgeons. Rev Col Bras Cir. 2021;48:e20213123. https://doi.org/10.1590/0100-6991e-20213123
15. Cunico C, Nasr A. Burnout e assédios na carreira cirúrgica: retirando a máscara e expondo as feridas. Rev Med (Paraná). 2020;78(1):119-22.
16. Grech E, Pace A, Attard Mallia T, Cuschieri S. Gender inequality and sexual harassment: a reality in medical school in 2022? A Malta single-centre study. Ethics Med Public Health. 2023;27:100888. https://doi.org/10.1016/j.jemep.2023.100888
17. Hawes AM, Gondy K. Sexual harassment in medical education: how we can do better. J Gen Intern Med. 2021;36(12):3841-3. https://doi.org/10.1007/s11606-021-06960-w
18. Maciel RH, Gonçalves RC. Pesquisando o assédio moral: a questão do método e a validação do Negative Acts Questionnaire (NAQ) para o Brasil [monografia]. 2008.
19. Sue DW. Microaggressions in everyday life: race, gender, and sexual orientation. 2nd ed. Hoboken (NJ): Wiley; 2020.
20. Low ZX, Yeo KA, Sharma VK, Leung GK, McIntyre RS, Guerrero A, et al. Prevalence of burnout in medical and surgical residents: a meta-analysis. Int J Environ Res Public Health. 2019;16(9):1479. https://doi.org/10.3390/ijerph16091479

21. Drossard S, Warnken I. Teaching medical students to navigate workplace harassment - preliminary experiences from a pilot workshop in Germany. BMC Med Educ. 2025;25(1):1251. https://doi.org/10.1186/s12909-025-07853-w

22. Ibrahim D, Riley R. Female medical students' experiences of sexism during clinical placements: a qualitative study. Healthcare (Basel). 2023;11(7):1002. https://doi.org/10.3390/healthcare11071002

23. Barnes KL, Dunivan G, Sussman AL, McGuire L, McKee R. Behind the mask: an exploratory assessment of female surgeons' experiences of gender bias. Acad Med. 2020;95(10):1529-38. https://doi.org/10.1097/ACM.0000000000003569

24. Gianakos AL, Freischlag JA, Mercurio AM, Haring RS, LaPorte DM, Mulcahey MK, et al. Bullying, discrimination, harassment, sexual harassment, and the fear of retaliation during surgical residency training: a systematic review. World J Surg. 2022;46(7):1587-99. https://doi.org/10.1007/s00268-021-06432-6

25. Fevre R, Robinson A, Jones T, Lewis D. Researching workplace bullying: the benefits of taking an integrated approach. Int J Soc Res Methodol. 2010;13(1):71-85. https://doi.org/10.1080/13645570802648671

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251821

Limitations of cardiac risk scores in elective non-cardiac surgeries: a Brazilian cohort

Fabio Figueiredo Costa[1], Liz Lustoza Brandão[2], Fernanda Hora Gomes de Sá[3], Paulo Átila Castro Carvalho de Jesus[4], João Vitor Souza Borges[5], Anna Claudia Monteiro Luz Santos[1], Jennifer do Carmo Souza Pinheiro[1], Lívia Brito Oliveira[6], Glicia Gleide Gonçalves Gama[7], Adriana Lopes Latado[8]*

SUMMARY

INTRODUCTION: Surgical procedures are associated with an excess of postoperative cardiovascular events. Risk scores are used for cardiovascular stratification but lack proper validation across diverse populations.

OBJECTIVE: The aim of this study was to assess the accuracy of the Revised Cardiac Risk Index, American College of Physicians score, and intuitive clinical judgment in predicting postoperative cardiovascular complications.

METHODS: In this prospective, longitudinal study, adults assessed for cardiovascular risk undergoing elective non-cardiac surgery were included. Perioperative clinical variables, Revised Cardiac Risk Index, American College of Physicians, and clinical judgment were evaluated, as well as in-hospital and 30-day postoperative outcomes. Data were analyzed using descriptive statistics, and score performance was assessed by C-statistics.

RESULTS: A total of 420 participants, with a mean age of 64.1±12.6 years, were studied, of which 47.1% were female and 71.2% were Black/Hispanic. Among the studied patients, 74.3% had hypertension, 32.9% were diabetics, 13.1% had heart failure, 11.2% had coronary artery disease, and 7.9% had stroke. Preoperative risk levels (low/medium/high) were as follows: Revised Cardiac Risk Index 88%/10.5%/1.4%; American College of Physicians 71.2%/27.4%/1.4%; and clinical judgment 64.5%/33.1%/2.4%. During hospitalization, there were 11 deaths, 3 myocardial infarctions, and 1 acute pulmonary edema. The risk scores showed poor discrimination for death and combined in-hospital cardiovascular outcomes, with area under the receiver operating characteristic curve <0.70. Cardiologist judgment performed slightly better than the scores. Results were similar for major and non-major surgeries, and when evaluating only patients aged 50 years or older.

CONCLUSION: Cardiac risk assessment demonstrated low accuracy in predicting in-hospital postoperative cardiovascular outcomes, regardless of surgical magnitude.

KEYWORDS: Risk assessment. General surgery. Cardiovascular diseases.

INTRODUCTION

Perioperative cardiac complications in non-cardiac surgeries account for a considerable absolute number of adverse events, given the high number of procedures performed annually in Brazil[1]. Risk stratification scores are applied in clinical practice to estimate the likelihood of major cardiovascular complications, cardiovascular death, and all-cause mortality[2-5]. These tools enhance patient care by providing evidence-based information to support clinical decision-making and mitigate adverse outcomes[1].

[1]Universidade Federal da Bahia, Hospital Universitário Professor Edgard Santos, Empresa Brasileira de Serviços Hospitalares, Cardiorespiratory System Unit – Salvador (BA), Brazil.
[2]Secretaria da Saúde do Estado da Bahia – Salvador (BA), Brazil.
[3]Universidade Estadual de Campinas – Salvador (BA), Brazil.
[4]Universidade Federal da Bahia – Salvador (BA), Brazil.
[5]Foundation José Silveira – Salvador (BA), Brazil.
[6]Universidade Federal da Bahia, Hospital Universitário Professor Edgard Santos, Empresa Brasileira de Serviços Hospitalares, Research and Technological Innovation Management Sector – Salvador (BA), Brazil.
[7]Universidade Federal da Bahia, Hospital Universitário Professor Edgard Santos, Empresa Brasileira de Serviços Hospitalares, Medical Clinic Unit – Salvador (BA), Brazil.
[8]Universidade Federal da Bahia, Hospital Universitário Professor Edgard Santos, Empresa Brasileira de Serviços Hospitalares, Faculdade de Medicina, Cardiorespiratory System Unit and Research and Technological Innovation Management Sector – Salvador (BA), Brazil.
*Corresponding author: adrianalatado@cardiol.br
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on October 09, 2025. Accepted on January 21, 2026.
Scientific Editor: Roseli Nomura

Subjective clinical judgment remains valuable in assessing perioperative cardiovascular risk; however, validated risk scores facilitate a more rational application of complementary tests and guide the requirement for cardiology support[1].

In Brazil, the surgical population is becoming older, while minimally invasive procedures with lower intrinsic risk have been increasingly adopted. Endoscopic procedures and low-complexity surgeries, such as cataract extraction, are also rising, but these scenarios are not assessed by traditional risk scores[1-4].

There are few scientific publications regarding the accuracy and validation of cardiac surgical risk assessment in non-cardiac procedures in low- and middle-income countries. Fatal and non-fatal postoperative complications, both hospital-acquired and long-term, can be elevated in this setting, especially for elderly patients[6,7]. In Brazil, the surgical population is aging, while minimally invasive procedures with lower intrinsic risk are increasingly being adopted. Endoscopic procedures and low-complexity surgeries, such as cataract extraction, are also on the rise, but these scenarios are not assessed using traditional risk scores[1-4].

This study aimed to compare the performance of two risk stratification tools commonly applied in Brazil—the Revised Cardiac Risk Index (RCRI) and the American College of Physicians (ACP) index[1]—with clinical judgment for predicting perioperative cardiovascular complications in patients undergoing non-cardiac surgery at a university hospital in Northeastern Brazil, stratified by surgical complexity.

METHODS

Design, sample, and eligibility

This study was conducted at a federal university hospital in Bahia, Brazil, a referral center for medium- and high-complexity care serving the entire state of Bahia. It was a prospective, observational, analytical cohort study including patients undergoing non-cardiac surgery at the institution.

Participants were recruited through a convenience sampling method. The sample comprised outpatients evaluated in a clinic dedicated to preoperative cardiac risk assessment and inpatients referred for preoperative cardiological evaluation between April 2019 and December 2023. All clinical assessments were performed by three experienced cardiologists, members of the research team and regular staff of the hospital.

Eligibility criteria involved patients aged ≥18 years, referred for cardiovascular risk assessment, and undergoing elective non-cardiac surgery, performed at the study hospital. Major surgeries included suprainguinal vascular, intra-abdominal, intrathoracic, and neurosurgical procedures.

Patients were excluded if they did not undergo surgery for any reason or if the surgery was performed in other health facilities. Cases of unobtainable 30-day postoperative outcome data (e.g., incorrect or incomplete contact information) were assessed for in-hospital outcomes only.

Study procedures

An interview was conducted for all participants. Data were collected through structured case report forms, performed during patient interviews and supplemented by reviewing medical records. Variables included demographic and clinical characteristics, current medications, laboratory and imaging findings, cardiac risk assessment by RCRI and ACP scores, subjective clinical judgment of the cardiologist, and clinical outcomes.

Postoperative outcomes were recorded during hospitalization and reassessed 30 days after surgery by telephone contact. When phone contact was unsuccessful, information was obtained from checking physicians or hospital records.

Cardiac risk scores

Both scores, ACP and RCRI scores, were calculated as recommended by the Perioperative Cardiovascular Assessment Guideline of the Brazilian Society of Cardiology, 2024[1]. The ACP index categorizes patients into low (<3%), intermediate (3–15%), or high (>15%) risk for postoperative coronary events[4,8], while RCRI was stratified into classes I (0.4% risk), II (0.9% risk), III (6.6% risk), and IV (11% risk)[2,8]. Major cardiac complications considered in the RCRI include myocardial infarction (MI), pulmonary edema, ventricular fibrillation or primary cardiac arrest, and complete heart block within 5 postoperative days[2,8].

Outcomes of interest

Study outcomes were aligned with the RCRI and ACP scores: fatal and non-fatal acute MI, acute pulmonary edema, ventricular fibrillation or primary cardiac arrest, and complete heart block. Outcomes were assessed up to postoperative day 5, with additional events recorded up to 30 days after surgery. The research team that evaluated outcomes was different from the professionals who performed the preoperative cardiac risk assessment, although formal data blinding strategies between the teams were not applied.

Statistical analysis

Quantitative variables were summarized using mean±standard deviation, or median±interquartile range, while categorical variables were described using proportion. Normality of continuous data was assessed using the Shapiro-Wilk test and distributional

characteristics (kurtosis and skewness). Discriminatory ability of the scores for predicting cardiovascular outcomes was evaluated using the C-statistic (area under the ROC curve), comparing the RCRI, ACP, and clinical judgment. Calibration was tested with the Hosmer-Lemeshow method. Surgical procedures were classified as major or non-major for subgroup analyses. Statistical analyses were performed using R software (R Core Team, Vienna, Austria), Windows version (open-source).

Ethical considerations

The study protocol was approved by the institutional Research Ethics Committee (approval no. 3.186.630; CAAE no. 05853319.2.0000.0049; March 8, 2019). The study complied with Brazilian National Health Council Resolution 466/2012 and its operational guidelines. Written informed consent was obtained from all participants.

RESULTS

A total of 420 patients undergoing elective non-cardiac surgery between April 2019 and December 2023 were included. Their mean age was 64.1±12.6 years; 47.1% were female and 71.2% self-identified as Black or mixed-race. Regarding comorbidities, 74.3% had hypertension, 32.9% diabetes mellitus, 45.7% dyslipidemia, and 13.1% prior heart failure. Additionally, 11.2% had a history of coronary artery disease and 7.9% had a previous stroke. Baseline demographic, clinical, and laboratory characteristics are shown in Table 1.

Concerning surgical procedures, 17.4% were classified as major surgeries. General abdominal surgery represented 22.3% of the cases, of which 7.8% were major. Besides, only 1.2% of the patients underwent intrathoracic procedures and 2.2% underwent neurosurgery. Suprainguinal vascular procedures were performed in 6.2% and infrainguinal vascular procedures in 10.3% of the patients. Carotid procedures accounted for 0.7%, intracerebral arterial surgeries for 1.9%, and arteriovenous fistula creation for 2.6%. Low-risk procedures such as endoscopies and ophthalmologic or dermatologic surgeries represented 33.4% within the sample.

Preoperative risk prediction by clinical scores

The predictive performance of both RCRI and ACP algorithms was evaluated for in-hospital and 30-day postoperative cardiovascular outcomes. Based on available data from 418 patients, risk stratification using RCRI showed that 368 patients (88.0%) were at low risk, 44 patients (10.5%) were at intermediate risk, and 1.4% were at high risk. In the same way, the ACP score (n=420) classified the majority of patients as low risk (71.2%), followed by 27.4% as intermediate risk and then 1.4% as high risk. Compared with the RCRI, subjective clinical judgment by cardiologists yielded higher proportions of patients categorized as intermediate risk (33.1%), with fewer classified as low risk (64.5%).

Stratification by surgical complexity showed that high-risk classification was observed at a higher rate among major surgeries versus non-major surgeries according to the RCRI (5 vs. 0.8%) and clinical judgment (4.7 vs. 2%), but not with the

Table 1. Baseline clinical, demographic, and laboratory characteristics of the sample.

Characteristic	N=420
Age	
Mean (SD)	64.1 (12.6)
Median (IQR)	65.5 (57–62)
Female sex, n (%)	198 (47.1)
Race/ethnicity, n (%)	
Black	151 (36.0)
Brown (mixed)	148 (35.2)
White	120 (28.6)
Comorbidities, n (%)	
Hypertension	312 (73.4)
Diabetes mellitus	138 (32.9)
Dyslipidemia	192 (45.7)
Coronary artery disease	47 (11.2)
Heart failure	54 (13.1)
Stroke/TIA	33 (7.9)
Hemodynamic data, mean (SD)	
Heart rate	77 (12.7)
Systolic blood pressure	138 (20.3)
Diastolic blood pressure	81.4 (11.3)
BMI, mean (SD)	27.5 (5.92)
Laboratory tests	
Urea, mg/dL[a]	40.4 (25.4)
Creatinine, mg/dL[a]	1.3 (1.4)
Sodium, mEq/L[a]	139 (3.7)
Potassium, mEq/L[a]	4.5 (0.6)
Glucose, mg/dL	
Mean (SD)	112 (45.5)
Median (IQR)	100 (88–120)
Hemoglobin, g/dL[a]	13.1 (2.0)
Hematocrit, %[a]	39.2 (5.8)

SD: standard deviation; IQR: interquartile range; TIA: transient ischemic attack; BMI: body mass index; SBP: systolic blood pressure; DBP: diastolic blood pressure. [a]Mean (SD).

ACP algorithm, which showed similar distributions between major and non-major surgeries (1.6 vs. 1.4%, respectively).

In-hospital and 30-day cardiovascular outcomes

Clinical outcomes included death, non-fatal, and fatal MI, pulmonary edema, complete atrioventricular block, ventricular fibrillation, ventricular tachycardia, and heart failure. During hospitalization, there were 11 deaths (2.6%), 3 cases of non-fatal MI (0.7%), and 2 cases of pulmonary edema (0.7%), one of which occurred concomitantly with MI. One additional new-onset case of heart failure was observed after discharge within 30 days, with no further adverse events during this period.

Among the 15 in-hospital events, 6 occurred in patients undergoing major surgery and 9 in those undergoing non-major surgery. According to the RCRI, adverse events occurred in 2.7% of low-risk and 11.4% of intermediate-risk patients. For the ACP score, corresponding rates were 3.3 and 4.3%, respectively. No events were recorded among patients classified as high risk by either score. Clinical judgment classified low-risk patients with 1.8% event rates and intermediate-risk patients with 7.2%, while no events were observed among those deemed high risk.

In major surgeries (n=64), six in-hospital cardiovascular events were observed. Event rates among patients classified as low risk were 8.3% (RCRI), 8.7% (ACP), and 6.7% (clinical judgment). Among intermediate-risk patients, event rates were 15.4% (RCRI), 11.8% (ACP), and 12.9% (clinical judgment).

In non-major procedures, event rates among low-risk patients were 1.9% (RCRI), 2.4% (ACP), and 1.2% (clinical judgment). For intermediate-risk patients, rates were 9.7, 3.1, and 5.6%, respectively.

Accuracy of risk scores in predicting mortality and postoperative cardiovascular outcomes

For the RCRI, combined endpoints included death, heart failure, ventricular arrhythmia (VF/VT), pulmonary edema, and MI. For the ACP score, endpoints were death and non-fatal MI. As only one cardiovascular event occurred after hospital discharge, 30-day outcomes were not included in discrimination analyses.

As shown in Figure 1, both RCRI and ACP demonstrated poor discriminatory ability (area under the curve [AUC]: 0.62 and 0.54, respectively), with the ACP score performing worse. Cardiologists' clinical judgment showed slightly better performance, though still limited, with an AUC of 0.68 for postoperative cardiovascular events (Figure 1).

Additional ROC analyses stratified by surgical complexity (major vs. non-major) yielded similar findings (Figure 2). Sensitivity analyses restricted to patients aged ≥50 years—closer to the populations used in original score validations—also demonstrated poor discrimination across all tools, with AUCs ranging from 0.51 to 0.66.

DISCUSSION

This study evaluated a sample of patients undergoing non-cardiac surgery, aiming to assess the accuracy of two widely used cardiac risk scores in both national and international clinical practice, and the risk assessment provided by experienced cardiologists. The results demonstrated that both the RCRI and ACP scores had low prediction for fatal and non-fatal in-hospital cardiovascular events, with areas under the ROC curve <0.70. The incidence of in-hospital outcomes was 4%, with 2.6% mortality, and only one case of new-onset heart failure occurred after hospital discharge within 30 days post-surgery. These rates were lower than those reported by Heinisch et al., who observed cardiovascular complications in 16% of their sample, with no cardiovascular deaths reported[8]. This difference may be explained by the type of surgeries performed (more than 60% major procedures in Heinisch's results, compared with 17.4% in ours), baseline population characteristics, comorbidities, treatment patterns

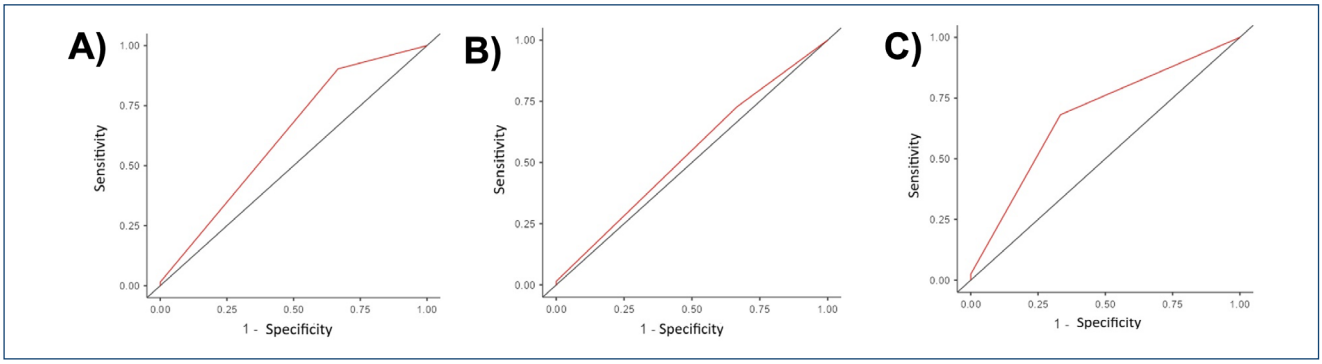


Figure 1. Areas under the receiver operating characteristic curve for the association between risk score levels and postoperative in-hospital clinical outcomes in patients undergoing elective non-cardiac surgery. (A) Revised Cardiac Risk Index score=0.62; (B) American College of Physicians score=0.54; (C) clinical judgment=0.68.

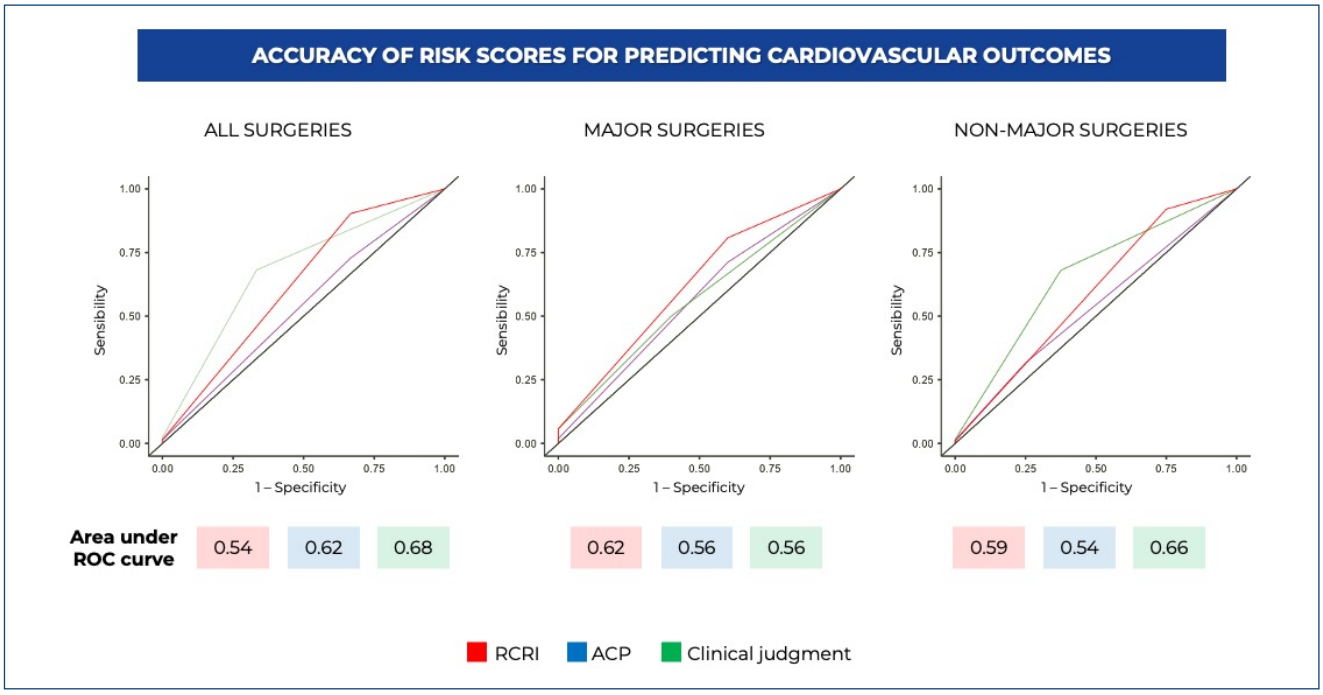


Figure 2. Comparison of the performance of Revised Cardiac Risk Index and American College of Physicians scores, and the cardiologist's clinical judgment, in the total sample, in major surgery and in non-major surgery.

(including greater statin use in our cohort), less invasive surgical techniques, and safer contemporary anesthetic protocols.

In our study, most patients were classified as low risk by preoperative risk scores, yet cardiologists' subjective assessment led to a significant increase in intermediate-risk classification, particularly when compared with the RCRI score. Surgical magnitude, especially major procedures, appeared to influence the perceived risk level by subjective analysis. Clinical judgment, although susceptible to both intuitive and analytical biases, showed slightly higher accuracy than the risk scores, albeit with unsatisfactory performance (ROC curve=0.68). Devereaux et al. similarly highlighted these discrepancies, reporting weak agreement (kappa=0.38) between physicians' subjective estimates and validated risk scores for non-cardiac surgery[9]. In their study, physicians overestimated risk in 16% of patients—likely generating unnecessary costs from additional diagnostic testing—and underestimated risk in 13%, potentially leading to insufficient perioperative evaluation and increased adverse outcomes. Although probabilistic risk models are not perfect, they are superior to the subjective judgments of physicians because human judgment is based on experience and intuition and is therefore susceptible to cognitive biases. These mental shortcuts, while useful in daily decision-making, can lead to distortions and errors in perioperative risk assessment[10].

The RCRI was originally derived from a prospective single-center cohort of 2,893 patients aged ≥50 years undergoing elective major non-cardiac surgery, with no inclusion of low-complexity procedures—differing substantially from our sample[2]. The RCRI showed moderate accuracy in its original validation cohort (ROC area=0.812)[2]. However, subsequent studies applying the RCRI in heterogeneous populations reported lower predictive values, with ROC area varying between 0.62 and 0.69[11]. These values, much lower than those in the original cohort, align more closely with our findings. Several factors may explain this difference. First, our study included many patients undergoing low-risk procedures, in contrast to the original RCRI cohort. Second, we included patients aged ≥18 years, whereas the RCRI was derived only from patients ≥50 years. In this regard, however, even in sub-analyses of patients over 50 years of age, the RCRI score did not prove to be accurate. Finally, advances over the past 25 years—including widespread laparoscopic and endovascular techniques and improved anesthesia—differ significantly from the context of the original RCRI derivation cohort, which primarily underwent traditional open surgeries.

Heinisch et al. also reported poor performance of preoperative scores—ASA, Detsky, Goldman, and Larsen—with ROC areas of 0.48, 0.38, 0.48, and 0.49, respectively, inferring that cardiac risk indices have limited value for predicting perioperative cardiac events, despite their widespread clinical use[12]. Applying these scores to populations markedly different from their derivation cohorts may partly explain these results. Additionally, the lack of systematic postoperative cardiovascular monitoring—such as serial electrocardiograms and biomarkers (troponin or creatine kinase [CK]-MB)—as well

as changes in the definition of acute MI over time, may also play a significant role in differences between study results.

Cardiac complications in the perioperative setting of non-cardiac surgery remain a major concern. The use of traditional risk scores—such as RCRI and modified ACP—or cardiologists' subjective assessments demonstrates limited accuracy in populations dominated by low-risk procedures. Careful interpretation of these results is required to avert unnecessary use of medical resources and the unwarranted postponement or cancellation of essential surgical procedures. The recently proposed Beirut score offers ease of application and requires minimal diagnostic resources, with greater representation of low-risk procedures in its derivation cohort. However, like other perioperative scores, it originated from a single US center, which raises concerns about external validity and applicability in other settings, including Brazil[5].

In current clinical practice, patients deemed "low risk" by established scores may still face a significant risk of adverse consequences not captured by these models[2,5,13,14]. Several comorbidities identified during preoperative assessment—such as obesity, smoking, or chronic lung disease—are not included in existing scores, despite their potential impact on both short- and long-term perioperative outcomes[14,15].

New algorithms are currently being tested for predicting adverse outcomes in different clinical and surgical scenarios, some with the addition of machine learning technology, which have shown promise[16,17]. The incorporation of new analytical models, sometimes in addition to existing predictive tools, is important to update data on cohorts from different clinical and surgical contexts and to encompass more complex innovative technologies.

This study has limitations. First, the research took place in a single center, which may limit the generalizability of the findings. The lack of complete blinding of the evaluators in assessing hospital outcomes has the potential for bias, even though the verification of postoperative events was carried out by a research team different from the cardiologists who performed the preoperative risk assessment. The inferential statistical analyses were exploratory because a priori sample size was not calculated. On the other hand, our findings provide real-world clinical data from a highly heterogeneous population. Understanding this reality supports the implementation of strategies to improve perioperative care, optimize resource utilization, and reduce unnecessary surgical cancellations or delays. These improvements may ultimately have a positive economic impact, particularly in low- and middle-income countries such as Brazil.

In conclusion, this study found that preoperative cardiac risk assessment for predominantly low-complexity non-cardiac surgeries, whether based on ACP, RCRI, or expert clinical judgment, showed poor accuracy in predicting major fatal and non-fatal postoperative cardiovascular events, particularly in-hospital outcomes. The development of new risk scores, likely by incorporating machine learning insights, is warranted to improve patient stratification, minimize unnecessary testing, and avoid inappropriate postponement or cancellation of essential surgical procedures.

AUTHORS' CONTRIBUTIONS

FFC: Conceptualization, Data curation, Writing – original draft, Writing – review & editing. **LLB:** Data curation, Writing – original draft, Writing – review & editing. **FHGS:** Data curation, Writing – original draft, Writing – review & editing. **PÁCCJ:** Data curation, Writing – original draft, Writing – review & editing. **JVSB:** Data curation, Writing – original draft, Writing – review & editing. **ACMLS:** Data curation, Writing – original draft, Writing – review & editing. **JCSP:** Data curation, Writing – original draft, Writing – review & editing. **LBO:** Data curation, Formal Analysis, Writing – original draft, Writing – review & editing. **GGGG:** Data curation, Writing – original draft, Writing – review & editing. **ALL:** Conceptualization, Data curation, Formal Analysis, Writing – original draft, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Gualandro DM, Fornari LS, Caramelli B, Abizaid AAC, Gomes BR, Tavares CAM, et al. Guideline for perioperative cardiovascular evaluation of the Brazilian society of cardiology - 2024. Arq Bras Cardiol. 2024;121(9):e20240590. https://doi.org/10.36660/abc.20240590
2. Lee TH, Marcantonio ER, Mangione CM, Thomas EJ, Polanczyk CA, Cook EF, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. Circulation. 1999;100(10):1043-9. https://doi.org/10.1161/01.cir.100.10.1043
3. Palda VA, Detsky AS. Perioperative assessment and management of risk from coronary artery disease. Ann Intern Med. 1997;127(4):313-28. https://doi.org/10.7326/0003-4819-127-4-199708150-00012
4. American College of Physicians. Guidelines for assessing and managing the perioperative risk from coronary artery disease associated with major noncardiac surgery. Ann Intern Med. 1997;127:309-12.
5. Dakik HA, Chehab O, Eldirani M, Sbeity E, Karam C, Abou Hassan O, et al. A new index for pre-operative cardiovascular evaluation. J Am Coll Cardiol. 2019;73(24):3067-78. https://doi.org/10.1016/j.jacc.2019.04.023

6. Çiçek V, Cinar T, Hayiroglu MI, Kılıç Ş, Keser N, Uzun M, et al. Preoperative cardiac risk factors associated with in-hospital mortality in elderly patients without heart failure undergoing hip fracture surgery: a single-centre study. Postgrad Med J. 2021;97(1153):701-5. https://doi.org/10.1136/postgradmedj-2020-138679
7. Orhan AL, Çınar T, Hayıroğlu Mİ, Çiçek V, Selçuk M, Doğan S, et al. Atrial fibrillation as a preoperative risk factor predicts long-term mortality in elderly patients without heart failure and undergoing hip fracture surgery. Rev Assoc Med Bras (1992). 2021;67(11):1633-8. https://doi.org/10.1590/1806-9282.20210686
8. Feitosa-Filho GS, Loureiro BM, Nascimento Jdos S. Agreement between three perioperative risk scores. Rev Assoc Med Bras (1992). 2016;62(3):276-9. https://doi.org/10.1590/1806-9282.62.03.276
9. Devereaux PJ, Ghali WA, Gibson NE, Skjodt NM, Ford DC, Quan H, et al. Physician estimates of perioperative cardiac risk in patients undergoing noncardiac surgery. Arch Intern Med. 1999;159(7):713-7. https://doi.org/10.1001/archinte.159.7.713
10. Croskerry P. From mindless to mindful practice--cognitive bias and clinical decision making. N Engl J Med. 2013;368(26):2445-8. https://doi.org/10.1056/NEJMp1303712
11. Ford MK, Beattie WS, Wijeysundera DN. Systematic review: prediction of perioperative cardiac complications and mortality by the revised cardiac risk index. Ann Intern Med. 2010;152(1):26-35. https://doi.org/10.7326/0003-4819-152-1-201001050-00007
12. Heinisch RH, Barbieri CF, Nunes Filho JR, Oliveira GL, Heinisch LM. Prospective assessment of different indices of cardiac risk for patients undergoing noncardiac surgeries. Arq Bras Cardiol. 2002;79(4):327-38. https://doi.org/10.1590/s0066-782x2002001300001
13. Eagle KA, McKay RE. Pre-operative risk prediction: will better tools produce better outcomes? J Am Coll Cardiol. 2019;73(24):3079-81. https://doi.org/10.1016/j.jacc.2019.04.024
14. Correll DJ, Bader AM, Hull MW, Hsu C, Tsen LC, Hepner DL. Value of preoperative clinic visits in identifying issues with potential impact on operating room efficiency. Anesthesiology. 2006;105(6):1254-9;discussion 6A. https://doi.org/10.1097/00000542-200612000-00026
15. Gualandro DM, Fornari LS, Caramelli B, Abizaid A, Gomes BR, Tavares CAM, et al. Diretriz de avaliação cardiovascular perioperatória da sociedade Brasileira de cardiologia – 2024. Arq Bras Cardiol. 2024;121:e20240590.
16. Cicek V, Babaoglu M, Saylik F, Yavuz S, Mazlum AF, Genc MS, et al. A new risk prediction model for the assessment of myocardial injury in elderly patients undergoing non-elective surgery. J Cardiovasc Dev Dis. 2024;12(1):6. https://doi.org/10.3390/jcdd12010006
17. Cicek V, Orhan AL, Saylik F, Sharma V, Tur Y, Erdem A, et al. Predicting short-term mortality in patients with acute pulmonary embolism with deep learning. Circ J. 2025;89(5):602-11. https://doi.org/10.1253/circj.CJ-24-0630

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251838

Epidemiological profile of pregnant women undergoing first-trimester preeclampsia screening at a tertiary reference center in Brazil

Vivianne Netto Alves dos Reis[1], Edward Araujo Júnior[2,3]*, Beatriz Bussi Rosolen[3], Gustavo Yano Callado[4], Sandra Rejane Silva Herbst[1], Ingrid Schwach[1,2], Carolina Leite Drummond[1], Giselle Darahem Tedesco[1]

SUMMARY

OBJECTIVE: The aim of this study was to evaluate the epidemiological profile of first-trimester pregnant women attending a tertiary reference center in São Paulo, Brazil, and to assess its association with the development of preeclampsia.

METHODS: This study involved retrospective and prospective cohorts of pregnant women undergoing first-trimester ultrasound for aneuploidy screening between 11 and 13+6 weeks' gestation from 2020 to 2023. Maternal, obstetric, and clinical characteristics, including mean arterial pressure, were collected via questionnaire and clinical measurements. Statistical analyses included descriptive statistics and stepwise forward logistic regression to identify factors associated with preeclampsia development.

RESULTS: A total of 104 pregnant women were included, with a mean age of 30.9 years. Most were White (48.3%), non-smokers (86.5%), overweight (mean body mass index: 28.5 kg/m2), and had spontaneous pregnancies (100%). Nulliparous and primiparous women accounted for 71.9% of the sample. Low risk for preeclampsia was observed in 84.3% and high risk in 15.7%. History of previous preeclampsia (OR 46.0, 95%CI 3.2–654.2), chronic arterial hypertension (OR 11.2, 95%CI 1.9–64.6), and increased first-trimester mean arterial pressure (OR 1.2, 95%CI 1.1–1.4) were independently associated with higher preeclampsia risk. Other variables, including race, parity, body mass index, and uterine artery Doppler indices, were not significant predictors.

CONCLUSION: Prior history of preeclampsia, chronic hypertension, and elevated first-trimester mean arterial pressure were key predictors of preeclampsia development. These findings support the importance of first-trimester risk stratification using maternal history and mean arterial pressure measurement to guide early preventive interventions, particularly in settings where biochemical markers are not routinely available.

KEYWORDS: Preeclampsia. First trimester. Mean arterial pressure. Risk factors. Maternal health.

INTRODUCTION

The prevalence of preeclampsia (PE) is higher in regions with lower socioeconomic development, which account for approximately 99% of cases worldwide. In Brazil, PE affects 1.5–7% of pregnancies, with 0.6% progressing to eclampsia, although these estimates likely vary by region and may be underestimated[1,2]. PE is a multisystem disorder associated with long-term maternal morbidity, including increased risks of stroke, cardiovascular disease, and diabetes, as well as adverse neonatal outcomes primarily related to prematurity, such as perinatal mortality and impaired neurodevelopment[3].

According to the International Society for the Study of Hypertension in Pregnancy (ISSHP), PE is defined by new-onset hypertension (≥140/90 mmHg) after 20 weeks of gestation in previously normotensive women, accompanied by proteinuria, maternal organ dysfunction, and/or placental involvement[2]. PE may be classified as preterm (<37+0 weeks) or term (≥37+0 weeks), while pathophysiological studies often distinguish early-onset (<34 weeks) from late-onset disease.

First-trimester screening is central to PE prevention, as interventions such as acetylsalicylic acid and calcium supplementation reduce the incidence of preterm PE[4]. The Fetal Medicine Foundation (FMF) model, validated internationally, combines maternal characteristics, medical and obstetric history, biochemical markers, and biophysical parameters, including mean arterial pressure (MAP) and uterine artery Doppler, to identify women at increased risk[1,5]. In Brazil, the Brazilian Federation of Gynecology and Obstetrics Associations (FEBRASGO)

[1]Faculdade de Ciências Médicas da Santa Casa de São Paulo, Department of Obstetrics and Gynecology – São Paulo (SP), Brazil.
[2]Universidade Federal de São Paulo, Escola Paulista de Medicina, Department of Obstetrics – São Paulo (SP), Brazil.
[3]Universidade Municipal de São Caetano do Sul, Discipline of Women's Health – São Caetano do Sul (SP), Brazil.
[4]Hospital Israelita Albert Einstein, Albert Einstein Israelite Faculty of Health Sciences – São Paulo (SP), Brazil.
*Corresponding author: araujojred@terra.com.br
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on October 13, 2025. Accepted on December 22, 2025.
Scientific Editor: Roseli Nomura

guidelines emphasize universal screening based on medical history and MAP, tailored to local resource availability[6].

Given the substantial maternal and fetal burden of PE and the potential for early prevention, this study aimed to characterize the epidemiological profile of first-trimester pregnant women attending a tertiary reference center in São Paulo, Brazil, and to assess its association with the subsequent development of PE.

METHODS

This was a descriptive, observational study involving retrospective and prospective cohorts of pregnant women attending the Fetal Medicine Unit, Department of Obstetrics and Gynecology, Santa Casa Hospital, São Paulo, Brazil, between 2020 and 2023. The inclusion criterion was singleton pregnancies undergoing first-trimester ultrasound for aneuploidy screening (gestational age between 11 weeks and 13 weeks+6 days). Exclusion criteria included multiple pregnancies, fetal malformations, Müllerian anomalies, and miscarriages.

PE risk was calculated using the FMF first-trimester algorithm. Women were classified as high risk if the estimated risk for preterm PE (<37 weeks) was ≥1:100, and low risk if <1:100. MAP was measured using a validated automated oscillometric device, following FMF recommendations. After ≥5 min of rest in the seated position, two measurements were obtained from each arm with the appropriate cuff size and arms supported at heart level. The average of the four readings was used to calculate MAP.

Data were collected through a questionnaire administered at the time of the ultrasound examination, including socioeconomic variables (age, race, and smoking status); obstetric variables (gestational age determined by crown–rump length, parity, type of delivery, and prematurity); clinical variables (gestational diabetes mellitus, previous gestational diabetes, chronic arterial hypertension, systemic lupus erythematosus, antiphospholipid syndrome, history of PE in a previous pregnancy, maternal family history of PE, and history of previous small-for-gestational-age fetus); maternal weight, height, and body mass index (BMI); and combined MAP.

Statistical analysis was performed using SPSS version 25.0 (SPSS Inc., Chicago, IL, USA). For descriptive analysis, absolute (n) and relative (%) frequencies were calculated for qualitative variables, and summary measures (mean, standard deviation, median, minimum, and maximum) were calculated for quantitative variables. Chi-square or Fisher's exact tests were applied to qualitative predictor variables when appropriate, and statistically significant associations were expressed as odds ratios (ORs) with 95%CIs. For quantitative predictors, the Student's t-test or Mann-Whitney test was used according to data distribution. A significance level of 5% ($p<0.05$) was adopted for all tests.

For multivariate analysis, predictors with $p<0.20$ in univariate analysis were preselected, followed by stepwise forward logistic regression to identify clinically significant variables associated with the development of PE. Among these, chronic arterial hypertension, combined MAP, and a history of PE in a previous pregnancy were identified.

Normality was assessed using the Shapiro-Wilk test. Parametric or nonparametric tests were applied accordingly. MAP was analyzed as a continuous variable. Missing data were minimal (<5%) and handled by complete-case analysis.

RESULTS

In our study, 104 pregnant women who underwent a first-trimester scan for aneuploidy screening with concomitant risk calculation for PE were evaluated. The mean maternal age was 30.9 years, with most participants being White (48; 41.6%) and non-smokers (91; 87.5%). The mean BMI was 28.5 kg/m^2 (range: 16.6–48.6 kg/m^2), indicating that most women were overweight, and all pregnancies were spontaneous (104; 100%). The most frequent gestational age at examination was 12–13 weeks (72; 69.7%). Regarding parity, nulliparous and primiparous women together accounted almost all sample (37 nulliparas; 66 primiparas; total 103; 99.0%). Most women had no previous history of PE (98; 94.2%), chronic arterial hypertension (85; 81.7%), gestational diabetes mellitus (98; 94.2%), preexisting diabetes mellitus (96; 92.3%), or maternal history of PE (95; 91.3%) (Table 1).

When these parameters were evaluated, we found a low-risk rate for PE of 84.3% and a high-risk rate of 15.7%, although without statistical significance. Among the factors identified as being associated with a higher risk of PE development, a history of previous PE showed the strongest association ($p=0.005$), with an OR of 46.0 and a 95%CI 3.2–654.2, suggesting that pregnant women with this antecedent have an approximately 46-fold higher risk of developing PE in the current pregnancy. The presence of chronic arterial hypertension also demonstrated a strong association ($p=0.007$), with an OR 11.2 and a 95%CI 1.9–64.6. Additionally, an increase in combined MAP was associated with a progressive risk of PE ($p=0.006$), with an OR 1.2 and a 95%CI 1.1–1.4, reinforcing the importance of MAP measurement as an early screening tool (Table 2). Other analyzed variables, such as race, parity, BMI, weight, height, and mean uterine artery pulsatility index (PI mean), did not show statistical significance (Table 3).

Table 1. Maternal and gestational characteristics.

Maternal characteristics	n (%)
Maternal age (years)	
<35 years	70 (67.3)
≥35 years	34 (32.7)
Body mass index (kg/m²)	
Normal (18.5–25)	22 (21.15)
Overweight (≥25)	78 (75)
Underweight (<18.5)	4 (3.85)
Race	
White	48 (46.1)
Black	14 (13.5)
Mixed	42 (40.4)
Smoking	
Yes	13 (12.5)
No	91 (87.5)
Parity	
Nulliparous	37 (35.6)
Primiparous	66 (63.5)
Multiparous (≥2)	1 (1.0)
Chronic arterial hypertension	
Yes	19 (18.3)
No	85 (81.7)
Previous diabetes mellitus	
Yes	8 (7.7)
No	96 (92.3)
Gestational diabetes mellitus	
Yes	6 (5.8)
No	98 (94.2)
Previous preeclampsia	
Yes	6 (5.8)
No	98 (94.2)
Maternal history of preeclampsia	
Yes	9 (8.7)
No	95 (91.3)
Conception	
Spontaneous	104 (100)
Assisted	0 (0)

Table 2. Maternal characteristics and risk for early-onset preeclampsia.

	p-value	OR	95%CI
Chronic arterial hypertension	0.007	11.2	1.9–64.6
Combined mean arterial pressure	0.006	1.2	1.1–1.4
Previous preeclampsia	0.005	46.0	3.2–654.2

CI: confidence interval; OR: odds ratio.

Table 3. Quantitative variables and risk for early-onset preeclampsia.

Quantitative characteristics	Low risk mean (SD)	High risk mean (SD)	p-value
Height (m)	1.63 (0.07)	1.61 (0.06)	0.44
Weight (kg)	75 (18.0)	79 (17)	0.385
Mean arterial pressure (mmHg)	83.7 (75)	100 (12.0)	<0.001
Mean uterine pulsatility index of uterine arteries Doppler	21 (27)	19.2 (29.0)	0.673

SD: standard deviation.

DISCUSSION

In this first-trimester cohort of 104 pregnant women, the prevalence of high risk for PE was low (15.7%); however, three maternal factors were independently associated with disease development: previous PE, chronic arterial hypertension, and elevated MAP. A history of PE conferred the highest risk, with an approximately 46-fold increased likelihood of recurrence, followed by chronic arterial hypertension (11-fold increase). Incremental increases in MAP were also associated with progressively higher risk. In contrast, age, BMI, race, and parity were not significantly associated. These findings highlight the predictive value of clinical history and hemodynamic assessment using simple, reproducible parameters in early prenatal care.

Our findings that a prior history of PE is the strongest predictor of recurrence are consistent with previous cohort studies, which report recurrence rates ranging from 14 to 21%, with higher rates (up to 31%) for early-onset or severe cases[7-10]. The magnitude of association observed in our cohort (OR 46.0) aligns with prior reports indicating substantially increased adjusted relative risks or incidence rate ratios (aRRs/IRRs: 12.8–14.5) for recurrent PE, particularly for preterm disease or when combined with interpregnancy hypertension[7,9,10]. Additionally, comorbidities such as chronic hypertension, obesity, and diabetes further elevate the risk, supporting the multifactorial nature of recurrence. These results reinforce the clinical relevance of obtaining a detailed obstetric history during the first trimester and are in line with validated predictive models, including the FMF algorithm, which integrates maternal history, MAP, uterine artery Doppler, and placental growth factor (PIGF) to identify high-risk women early[7,10]. Importantly, although recurrence risk is high, most women with prior PE do not experience recurrence, and outcomes in subsequent pregnancies are generally favorable with individualized monitoring and prophylactic interventions, such as low-dose aspirin initiated before 16 weeks' gestation[10].

Mean uterine artery PI was not significantly associated with PE risk. The wide variability observed suggests limited

standalone predictive value in this cohort, consistent with evidence that uterine artery Doppler performs best when combined with clinical and biochemical markers.

The association of elevated first-trimester MAP with increased PE risk observed in our study is also supported by international evidence. MAP alone has demonstrated moderate predictive performance, with area under the curve (AUC) values ranging from 0.82 to 0.87 for overall and term PE, and up to 0.91 for later gestational ages in some populations[11,12]. However, predictive accuracy improves markedly when MAP is incorporated into multiparametric models alongside maternal characteristics, uterine artery Doppler, and angiogenic biomarkers such as PlGF, as in the FMF triple test, achieving detection rates of 75–90% for early-onset and preterm PE[13-15]. Our finding that MAP is independently associated with progressively increased risk (OR 1.2) underscore its utility as a simple, low-cost, and accessible early screening tool, particularly in settings where biochemical markers may not be routinely available[11,13-15]. These results corroborate previous studies demonstrating that first-trimester MAP measurement is a robust component of risk stratification and can guide timely prophylactic interventions.

Advanced modeling of uterine artery PI using fractional polynomials has improved risk estimation but remains insufficient alone. Increased PE risk has also been associated with spontaneous preterm and term birth in unaffected pregnancies, whereas iatrogenic preterm birth remains strongly linked to placental dysfunction in established disease[16,17].

This study has several limitations that should be acknowledged. First, the relatively small sample size and single-center design may limit the generalizability of the findings to other populations and healthcare settings. Second, as data collection was based partly on self-reported information, potential recall or reporting bias cannot be excluded. Third, biochemical markers such as PlGF or pregnancy-associated plasma protein-A (PAPP-A) were not incorporated into the analysis, which may have restricted the comprehensive assessment of the FMF screening model. Stepwise regression may introduce selection bias; thus, results should be interpreted cautiously. The absence of biochemical markers (PlGF, PAPP-A) likely reduced screening performance, as their inclusion improves detection rates in the FMF model. Nonetheless, this reflects real-world Brazilian settings where such markers are often unavailable. Despite these limitations, the study provides relevant epidemiological insights into first-trimester PE screening in a Brazilian tertiary care setting, contributing to the optimization of local prenatal care strategies.

CONCLUSION

Our findings support the use of maternal history and standardized MAP measurement as feasible first-trimester screening tools, particularly in Brazilian settings with limited access to biochemical markers. Although limited by its sample size, the findings highlight the importance of incorporating standardized first-trimester screening protocols into routine prenatal care to improve maternal and perinatal outcomes. Further multicenter studies with larger cohorts are needed to validate these results and to refine risk prediction strategies in the Brazilian population.

AUTHORS' CONTRIBUTIONS

VNAR: Conceptualization, Investigation. **EAJ:** Validation, Writing – original draft, Writing – review & editing. **BBR:** Resources, Writing – original draft. **GYC:** Data curation, Formal Analysis. **SRSH:** Software, Validation. **IS:** Methodology, Resources. **CLD:** Conceptualization, Methodology. **GDT:** Project administration, Supervision.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Poon LC, Shennan A, Hyett JA, Kapur A, Hadar E, Divakar H, et al. The International Federation of Gynecology and Obstetrics (FIGO) initiative on pre-eclampsia: a pragmatic guide for first-trimester screening and prevention. Int J Gynaecol Obstet. 2019;145(Suppl 1):1-33. https://doi.org/10.1002/ijgo.12802
2. Magee LA, Brown MA, Hall DR, Gupte S, Hennessy A, Karumanchi SA, et al. The 2021 International Society for the Study of Hypertension in Pregnancy classification, diagnosis & management recommendations for international practice. Pregnancy Hypertens. 2022;27:148-69. https://doi.org/10.1016/j.preghy.2021.09.008
3. Dimitriadis E, Rolnik DL, Zhou W, Estrada-Gutierrez G, Koga K, Francisco RPV, et al. Pre-eclampsia. Nat Rev Dis Primers. 2023;9(1):8. https://doi.org/10.1038/s41572-023-00417-6
4. Rolnik DL, Wright D, Poon LCY, Syngelaki A, O'Gorman N, Paco Matallana C, et al. ASPRE trial: performance of screening for preterm pre-eclampsia. Ultrasound Obstet Gynecol. 2017;50(4):492-5. https://doi.org/10.1002/uog.18816
5. Wright D, Wright A, Nicolaides KH. The competing risk approach for prediction of preeclampsia. Am J Obstet

Gynecol. 2020;223(1):12-23.e7. https://doi.org/10.1016/j.ajog.2019.11.1247

6. Peixoto-Filho FM, Costa FS, Kobayashi S, Beitune P, Garrido AG, Carmo AV, et al. Prediction and prevention of preeclampsia Number 1 – January 2023. Rev Bras Ginecol Obstet. 2023;45(1):49-54. https://doi.org/10.1055/s-0043-1763495
7. Nie Q, Zhou B, Wang Y, Ye M, Chen D, He F. Evaluation of outcomes and risk factors for recurrent preeclampsia in a subsequent pregnancy. Arch Gynecol Obstet. 2024;310(5):2487-95. https://doi.org/10.1007/s00404-024-07751-4
8. Ragnarsdóttir IB, Akhter T, Junus K, Lindström L, Lager S, Wikström AK. Does developing interpregnancy hypertension affect the recurrence risk of preeclampsia? A population-based cohort study. Am J Hypertens. 2024;37(7):523-30. https://doi.org/10.1093/ajh/hpae034
9. Lailler G, Grave C, Gabet A, Regnault N, Deneux-Tharaux C, Kretz S, et al. Recurrence of hypertensive disorders of pregnancy: results from a nationwide prospective cohort study (CONCEPTION). BJOG. 2023;130(8):902-12. https://doi.org/10.1111/1471-0528.17424
10. Chaemsaithong P, Sahota DS, Poon LC. First trimester preeclampsia screening and prediction. Am J Obstet Gynecol. 2022;226(2S):S1071-97.e2. https://doi.org/10.1016/j.ajog.2020.07.020
11. Zhu J, Zhang J, Syaza Razali N, Chern B, Tan KH. Mean arterial pressure for predicting preeclampsia in Asian women: a longitudinal cohort study. BMJ Open. 2021;11(8):e046161. https://doi.org/10.1136/bmjopen-2020-046161
12. Baschat AA, Magder LS, Doyle LE, Atlas RO, Jenkins CB, Blitzer MG. Prediction of preeclampsia utilizing the first trimester screening examination. Am J Obstet Gynecol. 2014;211(5):514.e1-7. https://doi.org/10.1016/j.ajog.2014.04.018
13. Rocha RS, Gurgel Alves JA, Bezerra Maia E Holanda Moura S, Araujo Júnior E, Martins WP, Vasconcelos CTM, et al. Comparison of three algorithms for prediction preeclampsia in the first trimester of pregnancy. Pregnancy Hypertens. 2017;10:113-7. https://doi.org/10.1016/j.preghy.2017.07.146
14. Serra B, Mendoza M, Scazzocchio E, Meler E, Nolla M, Sabrià E, et al. A new model for screening for early-onset preeclampsia. Am J Obstet Gynecol. 2020;222(6):608.e1-18. https://doi.org/10.1016/j.ajog.2020.01.020
15. Crovetto F, Figueras F, Triunfo S, Crispi F, Rodriguez-Sureda V, Dominguez C, et al. First trimester screening for early and late preeclampsia based on maternal characteristics, biophysical parameters, and angiogenic factors. Prenat Diagn. 2015;35(2):183-91. https://doi.org/10.1002/pd.4519
16. Cavoretto PI, Salmeri N, Candiani M, Farina A. Reference ranges of uterine artery pulsatility index from first to third trimester based on serial Doppler measurements: longitudinal cohort study. Ultrasound Obstet Gynecol. 2023;61(4):474-80. https://doi.org/10.1002/uog.26092
17. Cavoretto PI, Farina A, Salmeri N, Syngelaki A, Tan MY, Nicolaides KH. First trimester risk of preeclampsia and rate of spontaneous birth in patients without preeclampsia. Am J Obstet Gynecol. 2024;231(4):452.e1-7. https://doi.org/10.1016/j.ajog.2024.01.008

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251878

Cord insertion anomalies and adverse pregnancy outcomes: more than an incidental finding

Hasan Süt[1], Mustafa Koçar[1], Gülşah Aynaoğlu Yıldız[1]*

SUMMARY

OBJECTIVE: The aim of this study was to investigate the importance of routine evaluation of cord insertion during the prenatal period. To this end, the perinatal outcomes of marginal and velamentous cord insertions were compared to normal cord insertions.

METHODS: We conducted a retrospective cohort study of singleton pregnancies delivered after 24 weeks between 2016 and 2023, comparing marginal and velamentous cord insertions diagnosed on second-trimester ultrasound with normal cord insertions in terms of preterm birth and adverse perinatal outcomes, after excluding cases with fetal anomalies, placenta previa, placental invasion, vasa previa, or multiple gestations.

RESULTS: Among 6,675 singleton pregnancies, 367 (5.5%) had marginal cord insertion and 45 (0.7%) had velamentous cord insertion. Marginal cord insertion was associated with a nearly threefold higher risk of preterm birth before 37 weeks, while no significant differences were observed for other adverse outcomes. In contrast, velamentous cord insertion was linked to significantly increased risks of preterm birth, primary cesarean delivery, low Apgar scores, and neonatal intensive care unit admission compared with normal cord insertions.

CONCLUSION: Marginal and velamentous cord insertions showed associations with preterm birth and certain adverse perinatal outcomes. Routine evaluation of cord insertion may provide useful information for clinical follow-up.

Clinical Trial Registration: This study was retrospectively registered and authorized by the local Ethics Committee of Ankara University (clinical trial date 19.08.2025; number: 1942038).

KEYWORDS: Pregnancy outcomes. Ultrasound. Prenatal diagnosis. Premature birth. Umbilical cord abnormalities. Placenta abnormalities.

INTRODUCTION

The location of umbilical cord insertion (UCI) into the placenta is typically categorized as central, eccentric, marginal, or velamentous[1]. Among these, central and eccentric insertions are typically considered normal. Marginal cord insertion (MCI) is commonly defined as an insertion point located within 2 cm of the placental edge[2]. Velamentous cord insertion (VCI), on the other hand, occurs when the cord inserts into the fetal membranes instead of the placental disk, leaving the vessels unprotected as they traverse toward the placenta[3].

The UCI site can influence placental blood flow and fetal development[4,5]. Although several studies and meta-analyses have explored the association between abnormal UCI and adverse pregnancy outcomes such as fetal growth restriction, preterm birth, low Apgar scores, and increased rates of cesarean delivery, inconsistencies remain. This is particularly evident in second-trimester evaluations, as many studies are limited by small sample sizes, heterogeneity in study design, and variability in ultrasound timing and diagnostic criteria[6,7]. While the routine assessment of the UCI is not universally part of mid-trimester anomaly scans, current guidelines recommend reporting marginal or velamentous insertions when they are identified[8]. This emphasizes the ongoing debate and variability in clinical practice.

The present study aimed to evaluate the perinatal and neonatal outcomes associated with marginal and VCIs compared to normal cord insertions. We also sought to determine whether abnormal UCI is an independent risk factor for preterm delivery before 37 weeks of gestation.

METHODS

This retrospective cohort study was carried out at Ankara University Faculty of Medicine. The Ankara University Ethics Committee granted ethical approval on (12.08.2025) with approval number (1942038). Singleton pregnancies delivered at ≥24 weeks of gestation between February 2016 and December 2023 were reviewed. Pregnancies were excluded if a second-trimester ultrasound report was missing, the placental cord insertion site was not documented, or if structural or genetic fetal anomalies were present. Additional exclusion

[1]Ankara University, School of Medicine, Department of Obstetrics and Gynaecology – Ankara, Turkey.
*Corresponding author: gulsahayna@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on August 23, 2025. Accepted on January 07, 2026.
Scientific Editor: José Maria Soares Júnior

criteria included placenta previa, placental invasion anomalies, vasa previa, and multiple gestations. Following the application of these criteria, cases meeting the eligibility requirements were included in the analysis.

The placental cord insertion site was assessed using a Voluson E8 ultrasound machine (GE Healthcare, Zipf, Austria). Evaluations were performed by two experienced perinatologists, both with more than 10 years of expertise in fetal ultrasonography. In cases of diagnostic uncertainty, a third expert opinion was obtained. The assessment was conducted in accordance with the guidelines of the International Society of Ultrasound in Obstetrics and Gynecology (ISUOG)[8]. Cord insertion was evaluated using 2D grey-scale imaging and color or power Doppler in both the longitudinal and axial planes. MCI was defined as cord insertion within 2 cm of the placental edge. VCI was defined as the umbilical vessels inserting into the membranes before reaching the placental surface, with no protection from Wharton's jelly[9].

The following maternal characteristics were recorded: maternal age, gravidity, nulliparity, presence of comorbidities (hypertension, diabetes, preeclampsia, gestational hypertension, renal diseases, cardiac diseases, hypothyroidism, and hyperthyroidism), gestational age at diagnosis (in weeks), and history of uterine surgery. Delivery-related variables such as gestational age at birth, fetal sex, and mode of delivery were also documented.

The perinatal and neonatal outcomes that were assessed included pre-eclampsia, being small for gestational age (SGA), intrauterine growth restriction (IUGR), preterm birth before 37 or 34 weeks of gestation, stillbirth, emergency cesarean delivery (CD), primary CD, Apgar scores of less than 7 at 1 and 5 min, and admission to the neonatal intensive care unit (NICU). Prenatal outcomes in the marginal and VCI groups were compared with those in the normal cord insertion group. In addition, to identify independent factors associated with preterm birth before 37 weeks of gestation, univariate and multivariate logistic regression analyses were performed including maternal age,maternal comorbidity status, SGA, IUGR, type of cord insertion, CD, and preeclampsia as covariates. Maternal comorbidity was defined as the presence of at least one of the following conditions: hypertension, diabetes mellitus, preeclampsia, gestational hypertension, renal disease, cardiac disease, hypothyroidism, or hyperthyroidism.

Statistical analysis

All statistical analyses were performed using IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp., Armonk, NY, USA). Continuous variables were assessed for normality using the Kolmogorov-Smirnov test. Data were presented as mean ± standard deviation (SD) for normally distributed variables, and as median (interquartile range) for non-normally distributed variables. Categorical variables were expressed as numbers and percentages.

For comparisons between groups, the chi-square test or Fisher's exact test was used for categorical variables, and the independent samples t-test or Mann-Whitney U test was applied for continuous variables, as appropriate.

To identify factors associated with preterm birth before 37 weeks of gestation, univariate logistic regression analyses were first performed. Variables with $p<0.20$ in univariate analysis were included in a multivariate logistic regression model. Results were reported as odds ratios (ORs) with 95%CIs. A $p<0.05$ was considered statistically significant.

RESULTS

Between February 2016 and December 2023, data from a total of 7,980 deliveries at our clinic were reviewed. A total of 828 patients were excluded due to the absence of a second-trimester ultrasound or undocumented cord insertion site. Additionally, 84 cases with structural or genetic anomalies, 38 cases with placenta previa or placental invasion abnormalities, four cases with vasa previa, and 425 multiple pregnancies were excluded. Ultimately, 6,675 singleton pregnancies were included in the final analysis. Of these, 367 cases (5.5%) had MCI, 45 cases (0.7%) had VCI, and the remaining 6,263 cases (93.8%) had normal cord insertion.

The marginal and normal cord insertion groups exhibited similar characteristics regarding gravidity, nulliparity, gestational age at diagnosis, fetal gender, and mode of delivery. However, maternal age ($p<0.001$) and the presence of comorbidities ($p<0.001$) were significantly higher in the MCI group. Additionally, gestational age at delivery ($p<0.001$) and birthweight ($p=0.002$) were significantly lower in this group (Table 1).

In the comparison of perinatal and neonatal outcomes between the marginal and normal cord insertion groups, no significant differences were observed in the rates of preeclampsia, IUGR, SGA, stillbirth, emergency CD, low 1- and 5-min Apgar scores, or NICU admission. However, the MCI group had significantly higher rates of preterm birth before 37 weeks (OR 2.99; 95%CI 2.29–3.90; $p<0.001$) and primary CD (OR 1.31; 95%CI 1.02–1.68; $p=0.032$) (Table 2).

No significant differences were observed in the rates of preeclampsia, IUGR, SGA, stillbirth, or emergency CD when the perinatal outcomes of the 45 cases with VCI were compared to those with normal cord insertion. However, the VCI group had significantly higher rates of preterm birth <37 weeks (OR

3.58; 95%CI 1.8–7.1; p<0.001) and <34 weeks (OR 7.68; 95%CI 3.2–18.4; p<0.001), as well as higher rates of primary CD (OR 42; 95%CI 1.24–4.72; p=0.009), low 1-min (OR 3.67; 95%CI 1.62–8.3; p=0.002) and 5-min (OR 16.8; 95%CI 5.72–49.4; p<0.001) Apgar scores, as well as NICU admissions (OR 5.71; 95%CI 2.52–12.9; p<0.001) (Table 3).

To further evaluate the increased rate of preterm birth (<37 weeks) observed in both the MCI and VCI groups compared to the normal cord groups, a multivariable regression analysis was performed. Both MCI (OR 2.8; 95%CI 2.1–3.8; p<0.001) and VCI (OR 3.1; 95%CI 1.5–6.3; p=0.002) remained significant independent risk factors for preterm delivery in the adjusted model. In addition, preeclampsia, IUGR, CD, and the presence of maternal comorbidities were also independently associated with an increased risk of preterm birth (p<0.001, p<0.001, p<0.001, and p=0.013, respectively).

DISCUSSION

In this study, MCI was not linked to significant differences in adverse perinatal or neonatal outcomes such as preeclampsia, IUGR, SGA, stillbirth, emergency cesarean, low Apgar scores, or NICU admission. However, MCI was associated with a significantly higher risk of preterm birth before 37 weeks, occurring almost three times more often compared with normal cord insertion. In contrast, VCI was linked to markedly increased adverse outcomes, including higher rates of preterm birth (<37 and <34 weeks), low Apgar scores, and NICU admissions.

Table 1. Demographic features of marginal and normal cord insertion groups.

	Marginal cord insertion n=367	Normal cord insertion n=6,263	p-value
Maternal age (y)	29 (18–45)	28 (18–49)	**<0.001**
Gravidy	2 (1–8)	2 (1–10)	0.063
Nulliparity	132 (36%)	2,324 (37.1%)	.67
Comorbidity	203 (55.3%)	1,532 (24.5%)	**<0.001**
GA at diagnosis	21 (17–25)	21 (17–25)	0.092
The history of uterine surgery	90 (24.5%)	1,503 (24%)	0.80
GA at delivery (w)	38 (29–42)	39 (24–42)	**<0.001**
Gender (female)	176 (48.2%)	3,019 (48.2%)	0.94
Cesarean delivery	195 (53.1%)	3,025 (48.3%)	0.72
Birthweight (g)	3,190 (925–5,000)	3,250 (330–5,340)	**0.007**

The data are presented as median (minimum–maximum) or number (percentage). Statistically significant values (p<0.05) are in bold. GA: gestational age; g: gram; w: weeks; y: years.

Table 2. Perinatal and neonatal outcomes in marginal and normal cord insertion groups.

	Marginal cord insertion n=367	Normal cord insertion n=6,263	OR	95%CI	p-value
Preeclampsia	7 (1.9%)	161 (2.6%)	0.73	0.34–1.57	0.42
SGA	56 (15.3%)	789 (12.6%)	1.2	0.83–1.74	0.32
IUGR	23 (6.3%)	306 (4.9%)	1.1	0.63–1.9	0.73
Preterm delivery (<37 w)	78 (21.3%)	518 (8.3 %)	2.99	2.29–3.90	**<0.001**
Preterm delivery (<34 w)	5 (1.4%)	123 (2%)	0.68	0.28–1.69	0.42
Stillbirth	2 (0.5%)	18 (0.3%)	1.9	0.43–8.22	0.39
Emergency CD	15 (4.1%)	256 (4.1%)	1.0	0.58–1.7	0.99
Primary CD	108/276 (39.1%)	1,559/4,743 (32.9%)	1.31	1.02–1.68	**0.032**
1-min Apgar score of <7	15 (4.1%)	298 (4.8%)	0.71	0.38–1.31	0.28
5-min Apgar score of <7	4 (1.1%)	36 (0.6%)	2.64	0.8–8.75	0.11
NICU admission	18 (4.9%)	195 (3.1%)	1.6	0.98–2.64	0.059

The data are presented as number (percentage). Statistically significant values (p<0.05) are in bold. CD: cesarean delivery; CI: confidence interval; IUGR: intrauterine fetal growth restriction; NICU: neonatal intensive care unit; OR: odds ratio; SGA: small for gestational age.

Table 3. Perinatal and neonatal outcomes in velamentous and normal cord insertion groups.

	Velamentous cord insertion n=45	Normal cord insertion n=6,263	OR	95%CI	p-value
Preeclampsia	3 (6.7%)	161 (2.6%)	2.7	0.82–8.8	0.10
SGA	8 (17.8%)	789 (12.6%)	1.5	0.7–3.2	0.30
IUGR	4 (8.9%)	306 (4.9%)	1.9	0.67–5.32	0.22
Preterm delivery (<37 w)	11 (24.4%)	518 (8.3 %)	3.58	1.8–7.1	**<0.001**
Preterm delivery (<34 w)	6 (13.3%)	123 (2%)	7.68	3.2–18.4	**<0.001**
Stillbirth	0	18 (0.3%)	0	0	0.99
Emergency CD	4 (8.9%)	256 (4.1%)	2.28	0.81–6.42	0.11
Primary CD	19/35 (54.3%)	1,559/4,743 (32.9%)	2.42	1.24–4.72	**0.009**
1-min Apgar score of <7	7 (15.6)	298 (4.8%)	3.67	1.62–8.3	**0.002**
5-min Apgar score of <7	4 (8.9%)	36 (0.6%)	16.8	5.72–49.4	**<0.001**
NICU admission	7 (15.6%)	195 (3.1%)	5.71	2.52–12.9	**<0.001**

The data are presented as number (percentage). Statistically significant values (p<0.05) are in bold. CD: cesarean delivery; CI: confidence interval; IUGR: intrauterine fetal growth restriction; NICU: neonatal intensive care unit; OR: odds ratio; SGA: small for gestational age.

Overall, these findings suggest that assessing cord insertion may be justified as part of routine obstetric evaluation, considering its potential impact on pregnancy outcomes.

A meta-analysis investigating risk factors for MCI identified chronic hypertension, placenta previa, and nulliparity as significant contributors[10]. However, our study found that nulliparity was not a significant risk factor, whereas advanced maternal age and maternal comorbidities were.

Several retrospective studies have assessed the impact of MCI on perinatal outcomes. A 2021 study with 183 cases found no significant differences in intrapartum fetal intolerance, Apgar scores, birth weights, delivery complications, or NICU admissions[6]. Additionally, a 2022 study with 76 MCI cases (4.2%) and a 2025 study with 25 cases (5.6%) reported no statistically significant differences in adverse perinatal outcomes[1,11]. Conversely, a meta-analysis published in 2023, which combined data from 15 studies, found that MCI was associated with a higher risk of SGA, preeclampsia, stillbirth, preterm birth, and low birth weight[7]. Consistent with these results, the prevalence of MCI in this study was 5.5% (n=367), which is in line with previously reported rates. Notably, the MCI group showed a significantly increased risk of preterm birth and primary CD, while no significant differences were seen in other adverse perinatal or neonatal outcomes.

VCI has been more consistently associated with adverse perinatal outcomes compared to MCI, as shown in several studies[1,12,13]. A meta-analysis published in 2023 reported an overall incidence of 1.4% for VCI and demonstrated increased risks of SGA, preterm birth, and stillbirth; notably, when cases with vasa previa were excluded, the risks for SGA and stillbirth remained elevated, whereas the increased risk of preterm birth was no longer observed[14]. In our cohort, the incidence of VCI was 0.7%, which is slightly lower than that reported in the literature. Importantly, the VCI group in our study was characterized by higher rates of preterm delivery, primary cesarean section, low Apgar scores, and NICU admissions, further supporting the unfavorable perinatal impact of this placental abnormality.

One potential explanation for the variability in reported outcomes across studies, as well as a limitation of our own analysis, is related to the dynamic nature of placental cord insertions. A recent study demonstrated that cord insertion sites may "migrate" throughout pregnancy, with some marginal insertions resolving to normal positions later in gestation, while a small proportion may progress to velamentous insertion[4]. This phenomenon suggests that the timing of diagnosis and heterogeneity in ultrasound follow-up protocols could partly account for the inconsistent findings in the literature and should be considered when interpreting our results. In addition, the retrospective design of our study and the lack of systematic postnatal placental confirmation of cord insertion represent important limitations and may have influenced the precision of cord insertion classification in some cases. In another institution-based cross-sectional study, high umbilical artery PI values were found to be a reliable predictor of adverse perinatal outcomes in cases with cord insertion anomalies[15]. The absence of Doppler flow assessment in our cohort, therefore, represents an additional limitation of our study.

The strengths of our study include the large sample size, the single-center design, and the standardized evaluation of cord insertion by experienced perinatologists using ISUOG guidelines[8]. Furthermore, marginal and velamentous

insertions were analyzed separately, allowing a more precise comparison of outcomes. Another strength is the exclusion of high-risk conditions such as placenta previa, vasa previa, and multiple pregnancies, which helped reduce potential confounding. Limitations include the retrospective design, the absence of Doppler flow assessment, and the potential for cord insertion migration during pregnancy; additionally, the relatively small number of VCI cases may limit the precision of effect estimates and warrants cautious interpretation of subgroup analyses.

CONCLUSION

The current study shows that MCI is mainly linked to an increased risk of preterm birth. In contrast, VCI is associated with a broader range of adverse perinatal outcomes, including higher rates of preterm birth, low Apgar scores, and NICU admissions. These findings suggest that checking cord insertion during prenatal ultrasound could be an important part of perinatal assessment.

ACKNOWLEDGEMENTS

This article has been written for purely scientific purposes, and the authors have no conflict of interest with this article. No person/organization financially supports the study.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Our retrospective study included routine care procedures and received ethical approval from the local Ethics Committee of Ankara University. It was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from the participants.

PERMISSION TO REPRODUCE MATERIAL FROM OTHER SOURCES

All figures and materials are original and created by the authors. No permission is required.

AUTHORS' CONTRIBUTIONS

HS: Conceptualization, Writing – original draft, Writing – review & editing. **MK:** Conceptualization, Writing – review & editing. **GAY:** Conceptualization, Writing – original draft, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Pressman K, Markel L, Odibo A, Duncan JR. Perinatal outcomes based on placental cord insertion site. Am J Perinatol. 2026;43(2):199-203. https://doi.org/10.1055/a-2599-4764
2. Ismail KI, Hannigan A, O'Donoghue K, Cotter A. Abnormal placental cord insertion and adverse pregnancy outcomes: a systematic review and meta-analysis. Syst Rev. 2017;6(1):242. https://doi.org/10.1186/s13643-017-0641-1
3. Buchanan-Hughes A, Bobrowska A, Visintin C, Attilakos G, Marshall J. Velamentous cord insertion: results from a rapid review of incidence, risk factors, adverse outcomes and screening. Syst Rev. 2020;9(1):147. https://doi.org/10.1186/s13643-020-01355-0
4. Ward S, Sun Z, Maresse S. Placental cord insertion migration: implications for ultrasound documentation and follow-up of abnormal placental cord insertion site. Australas J Ultrasound Med. 2024;27(4):200-9. https://doi.org/10.1002/ajum.12399
5. Aragie H, Oumer M. Marginal cord insertion among singleton births at the University of Gondar Comprehensive Specialized Hospital, Northwest Ethiopia. BMC Pregnancy Childbirth. 2021;21(1):211. https://doi.org/10.1186/s12884-021-03703-x
6. Lutz AB, Young-Lin N, Leon-Martinez D, Bianco IC, Seckel E, Mrazek-Pugh B, et al. Measurement of marginal placental cord insertion by prenatal ultrasound was found not to be predictive of adverse perinatal outcomes. J Ultrasound Med. 2021;40(10):2079-86. https://doi.org/10.1002/jum.15586
7. Siargkas A, Tsakiridis I, Pachi C, Mamopoulos A, Athanasiadis A, Dagklis T. Impact of marginal cord insertion on perinatal outcomes: a systematic review and meta-analysis. Am J Obstet Gynecol MFM. 2023;5(4):100876. https://doi.org/10.1016/j.ajogmf.2023.100876
8. Salomon LJ, Alfirevic Z, Berghella V, Bilardo CM, Chalouhi GE, Silva Costa F, et al. ISUOG Practice Guidelines (updated): performance of the routine mid-trimester fetal ultrasound scan. Ultrasound Obstet Gynecol. 2022;59(6):840-56. https://doi.org/10.1002/uog.24888
9. Ward S, Sun Z, Maresse S. Current practice of placental cord insertion documentation in Australia - A sonographer survey. Australas J Ultrasound Med. 2023;26(3):157-68. https://doi.org/10.1002/ajum.12360
10. Siargkas A, Tsakiridis I, Gatsis A, Paco Matallana C, Gil MM, Chaveeva P, et al. Risk factors of marginal cord insertion in singleton pregnancies: a systematic review and meta-analysis. J Clin Med. 2024;13(23):7438. https://doi.org/10.3390/jcm13237438
11. Asoglu MR, Crimmins S, Kopelman JN, Turan OM, Goetzinger KR. Marginal placental cord insertion: the need for follow up? J Matern Fetal Neonatal Med. 2022;35(9):1629-35. https://doi.org/10.1080/14767058.2020.1763297

12. Visentin S, Londero AP, Santoro L, Pizzi S, Andolfatto M, Venturini M, et al. Abnormal umbilical cord insertions in singleton deliveries: placental histology and neonatal outcomes. J Clin Pathol. 2022;75(11):751-8. https://doi.org/10.1136/jclinpath-2020-207342

13. Wu XQ, Miao Y, Yang XF, Hong YQ, Wang LC, Chiu WH. Effect of abnormal placental cord insertion on hemodynamic change of umbilical cord in a tertiary center: a prospective cohort study. Postgrad Med J. 2025;101(1197):627-35. https://doi.org/10.1093/postmj/qgae193

14. Siargkas A, Tsakiridis I, Pachi C, Mamopoulos A, Athanasiadis A, Dagklis T. Impact of velamentous cord insertion on perinatal outcomes: a systematic review and meta-analysis. Am J Obstet Gynecol MFM. 2023;5(2):100812. https://doi.org/10.1016/j.ajogmf.2022.100812

15. Stapleton D, Darmonkow A, Ravi Chandran A, Milligan N, Saghian R, Shinar S, et al. Peripheral cord insertion is associated with adverse pregnancy outcome only when accompanied by clinically significant placental pathology. Ultrasound Obstet Gynecol. 2023;62(2):248-54. https://doi.org/10.1002/uog.26206

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251888

Evaluating chat generative pre-trained transformer responses to common patient questions on temporomandibular disorders

Bayram Sonmez Unuvar[1], Elif Esra Ozmen Unuvar[2]*

SUMMARY

OBJECTIVE: Temporomandibular disorders are among the most common causes of orofacial pain, often leading patients to seek information online. The increasing use of large language models such as chat generative pre-trained transformer in healthcare communication has raised questions about the reliability and readability of artificial intelligence-generated patient information. The aim of this study was to evaluate the accuracy, comprehensiveness, readability, and inter-rater reliability of chat generative pre-trained transformer-generated responses to common patient questions regarding temporomandibular disorders.

METHODS: ChatGPT (version 4.0) was prompted to generate 50 potential patient questions about temporomandibular disorders. Ten representative questions were selected and independently evaluated by five experts (two oral and maxillofacial surgeons, two physiotherapists, and one physical medicine specialist). Responses were rated using a four-point quality scale assessing accuracy and completeness. Readability was calculated using the Flesch-Kincaid method, and inter-rater reliability was assessed using the Intraclass Correlation Coefficient.

RESULTS: The responses demonstrated variable but generally acceptable quality. The overall Intraclass Correlation Coefficient value was 0.862, indicating good inter-rater agreement. Readability levels ranged from grade 6.2–10.7 (mean 8.0), corresponding to middle-to-high school comprehension. While most responses were rated satisfactory, several lacked sufficient clinical detail, particularly in differentiating professional consultation pathways.

CONCLUSION: chat generative pre-trained transformer provides moderately reliable and readable information about temporomandibular disorders, supporting its potential role in patient education. However, reliance on artificial intelligence-generated frequently asked questions introduces methodological limitations and authority bias. Future studies should incorporate real patient data and external fact-checking to enhance clinical relevance.

KEYWORDS: Temporomandibular disorders. Artificial intelligence. Patient education. Health communication. Large language models.

INTRODUCTION

Temporomandibular disorders (TMD) represent a group of musculoskeletal and neuromuscular conditions involving the temporomandibular joint (TMJ) and associated structures, often leading to pain, restricted jaw function, and impaired quality of life[1]. These disorders are multifactorial in origin, with contributing factors such as bruxism, trauma, muscle dysfunction, and psychosocial stress[2]. Due to their chronic and recurrent nature, TMDs can significantly affect daily activities including chewing, speaking, and social interactions, emphasizing the need for accessible and reliable patient education[3,4].

Patients increasingly turn to digital platforms to seek health-related information; however, the reliability and accuracy of online content are variable. In recent years, artificial intelligence (AI) tools based on large language models—such as chat generative pre-trained transformer (ChatGPT)—have been introduced as innovative means of delivering health information[5-7]. These systems can produce instant, user-friendly responses to complex questions, potentially supporting patient understanding and engagement. Despite this promise, concerns remain regarding the factual correctness, completeness, and consistency of AI-generated information[8,9].

Given the rapid adoption of ChatGPT in healthcare contexts, it is crucial to assess how accurately and comprehensively it conveys information about specific clinical conditions. Previous studies have examined its performance in fields such as orthopedics and oral surgery[10-12], but limited attention has been given to TMD, a condition requiring multidisciplinary understanding.

Importantly, the reliability of ChatGPT responses has often been tested using AI-generated or convenience-based question sets, introducing authority bias and limiting real-world applicability. To address this methodological limitation, the present

[1]KTO Karatay University, Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation – Konya, Türkiye.
[2]Karamanoglu Mehmetbey University, Department of Oral and Maxillofacial Surgery – Karaman, Türkiye.
*Corresponding author: elifesraozmen89@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on September 22, 2025. Accepted on December 24, 2025.
Scientific Editor: José Maria Soares Júnior

study focuses on evaluating ChatGPT's responses to commonly encountered patient questions about TMD.

The aim of this study is to evaluate the accuracy, comprehensiveness, readability, and inter-rater reliability of ChatGPT-generated responses to patient-oriented questions on TMD. By clarifying both the potential and limitations of AI-generated health information, this research aims to contribute to the evidence base for responsible and transparent use of large language models in patient education.

METHODS

Study design

This cross-sectional study was designed to evaluate the accuracy, comprehensiveness, readability, and inter-rater reliability of ChatGPT-generated responses to patient-oriented questions regarding TMD.

Question collection and selection

A preliminary Google Trends analysis indicated that patients often search for broad terms such as "jaw pain" more frequently than specific diagnostic terms like "TMJ pain." Guided by these findings, ChatGPT (version 4.0, OpenAI; accessed on [January 2025]) was prompted with: "List 50 common patient questions about TMD and their treatment."

From this initial list of 50 questions, the authors of the study independently reviewed all items to assess clarity, clinical relevance, and coverage of key patient concerns, including etiology, symptoms, treatment options, and self-care strategies. The selection process was exploratory and aimed to reflect common patient information needs rather than clinical prevalence. Minor discrepancies in initial preferences among the authors were resolved through consensus discussion, and full agreement was reached before finalizing the 10 questions. To reduce potential authority bias resulting from ChatGPT generating its own question set, this independent author review ensured diversity and clinical relevance. Importantly, the expert evaluators did not participate in the question selection process and were involved only in the subsequent independent evaluation of the finalized questions.

Data collection procedure

The 10 selected questions were entered sequentially into ChatGPT's online interface. Only the initial responses were recorded, without additional prompts or clarifications, to maintain standardization. Each answer was copied verbatim into a text document for analysis.

Expert evaluation

Five independent experts—two oral and maxillofacial surgeons, two physiotherapists, and one physical medicine and rehabilitation specialist—evaluated the responses using a four-point quality scale adapted from Mika et al.[13]. All evaluators were clinicians with a minimum of 5 years of professional experience in their respective fields and were actively involved in the routine diagnosis and management of patients with TMD. The multidisciplinary composition of the expert panel was intended to reflect real-world clinical practice in TMD management and to ensure that the evaluations were grounded in substantial domain-specific knowledge and clinical experience.

Experts were not blinded to the order of questions, as the primary objective was to assess the accuracy, completeness, and clarity of the content rather than to compare relative performance across items.

4 (Excellent): Comprehensive and fully accurate
3 (Satisfactory): Accurate with minimal missing details
2 (Moderate): Partially correct but incomplete
1 (Insufficient): Incorrect or lacking key information

Each expert rated the accuracy, completeness, and clarity of each response. Mean and range values were calculated for all items.

Readability assessment

The readability of ChatGPT's responses was assessed using the Flesch-Kincaid Reading Level via WordCalc software (https://www.wordcalc.com/readability/)[14]. Scores were reported as grade-level equivalents, with higher grades indicating more complex language.

Statistical analysis

All statistical analyses were performed using Statistical Package for the Social Sciences version 29.0 (IBM Corp., Armonk, NY, USA). Descriptive data were summarized as median (min–max). Inter-rater reliability among the five experts was calculated using the Intraclass Correlation Coefficient (ICC). ICC values were interpreted as: <0.50=poor, 0.50–0.75=moderate, 0.75–0.90=good, >0.90=excellent reliability[15].

RESULTS

Table 1 summarizes the readability levels and expert quality assessments of ChatGPT's responses to patient-oriented questions on TMD. The Flesch-Kincaid grade levels ranged from 6.2 to 10.7 (mean=8.0), corresponding to middle-to-high school readability.

Table 1. Readability (Flesch-Kincaid grade level) and quality ratings of chat generative pre-trained transformer's responses to common patient questions on temporomandibular disorders.

Questions	Flesch-Kincaid grade level	Quality of answer					ICC
		Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	
1. What are the most common causes of jaw pain?	7.7	2.00	2.00	2.00	2.00	2.00	0.862
2. Can jaw pain cause neck pain?	7.9	2.00	1.00	2.00	1.00	2.00	
3. Which specialist should I see for jaw pain?	7	1.00	1.00	1.00	1.00	1.00	
4. Should I apply heat or cold for jaw pain?	9.4	3.00	2.00	3.00	2.00	3.00	
5. Do massage or exercises help with jaw pain?	7.2	2.00	2.00	2.00	2.00	2.00	
6. What exercises are recommended for jaw pain?	6.2	2.00	1.00	2.00	1.00	2.00	
7. Is Botox treatment for jaw pain effective?	8.8	1.00	3.00	2.00	3.00	2.00	
8. Is a night splint effective for jaw pain?	7.6	2.00	2.00	2.00	2.00	2.00	
9. How long does it take to treat jaw pain?	8.7	3.00	3.00	3.00	2.00	3.00	
10. Do muscle relaxants relieve jaw pain?	10.7	2.00	2.00	2.00	2.00	2.00	

ICC indicates inter-rater reliability. Quality scores were based on a four-point scale (1=insufficient, 2=satisfactory, 3=good, 4=excellent) as described by Mika et al.[13]. ICC: Intraclass Correlation Coefficient (two-way random, absolute agreement). ICC=0.862 indicates good reliability.

The expert evaluations demonstrated consistent but variable quality across questions. Most responses were rated as satisfactory or moderately satisfactory, indicating that the content was generally accurate but occasionally lacked clinical depth or specific examples.

The five experts demonstrated high inter-rater reliability, with an ICC of 0.862 (see Table 1 caption).

Overall, the findings suggest that while ChatGPT provides readable and generally accurate information on TMD, certain responses remain too general or incomplete, particularly those requiring differential diagnosis or individualized treatment recommendations.

DISCUSSION

This study evaluated ChatGPT's responses to common patient questions about TMD, focusing on their accuracy, completeness, and readability. The findings indicate that ChatGPT provides moderately satisfactory and generally accurate information that may support patient education and health communication. However, its responses sometimes lack clinical specificity and contextual detail, which may limit their usefulness for complex or individualized cases. As AI tools such as ChatGPT become increasingly integrated into healthcare communication, evaluating their strengths and limitations remains essential for responsible implementation.

The readability analysis revealed Flesch-Kincaid grade levels ranging from 6.2 to 10.7 (mean=8.0), indicating that ChatGPT's responses were generally written at a middle-to-high school reading level. This suggests that most of the generated content is accessible to the general public, though certain responses may still require higher literacy skills for full comprehension. In comparison, previous studies assessing ChatGPT's outputs in surgical or orthopedic contexts—such as ulnar collateral ligament reconstruction—reported higher readability scores (mean=11.5), suggesting more complex text structures[10]. These findings imply that ChatGPT adapts its language complexity according to topic generality, producing more accessible content for broad conditions like TMD. Simplifying technical terminology and improving sentence clarity could further enhance patient understanding and health literacy.

In terms of response quality, most of ChatGPT's answers were rated as satisfactory or moderately satisfactory, indicating that the model generally provides accurate yet sometimes incomplete information. A few responses received lower scores, reflecting insufficient detail or lack of clinical specificity, particularly when the questions required nuanced diagnostic or therapeutic distinctions. Conversely, some items were rated higher for their clear structure and appropriate coverage of broad educational content, demonstrating ChatGPT's ability to organize and communicate information effectively. Minor differences among experts' ratings may be attributed to the inherent subjectivity of expert assessment and variation in clinical experience.

Previous research examining the accuracy and reliability of ChatGPT's responses to clinical and patient-centered questions has shown mixed results. Several studies report that the model can generate factually acceptable yet generalized information across various healthcare domains. For example, in evaluations

of ChatGPT's responses to frequently asked questions about anterior cruciate ligament reconstruction, most outputs were rated as satisfactory in content quality[12]. While studies on anterior cruciate ligament and ulnar collateral ligament injuries provide useful benchmarks for evaluating AI-generated patient information, TMD differ fundamentally in that they require a multidisciplinary approach involving dentistry, physiotherapy, and behavioral management rather than single-structure treatment strategies. However, for questions requiring procedure-specific or evidence-based detail, ChatGPT's responses tended to remain superficial, often lacking contextual nuance, source attribution, or clinical applicability[16]. These limitations highlight the model's dependence on its training corpus rather than real-time medical databases, which restricts its reliability for precise clinical decision support.

Comparable findings have been reported in studies evaluating ChatGPT's performance in other surgical and dental domains, including ulnar collateral ligament reconstruction, hip arthroscopy, and third molar surgery[10,11,17]. Across these contexts, the model has been found to generate adequate and well-structured responses to general patient inquiries but consistently falls short when compared to expert-generated explanations—particularly for questions requiring detailed clinical reasoning, procedural nuances, or evidence-based guidance. Furthermore, several investigations have emphasized the absence of source citation and transparency in ChatGPT's responses, noting that this omission may reduce users' trust and limit the model's educational reliability[11]. Addressing these issues by integrating verifiable references and contextual accuracy could substantially improve the safety and credibility of AI-based patient education tools.

In the field of dentistry, particularly within oral and maxillofacial surgery and dental radiology, recent studies have shown that ChatGPT can generate informative but generalized responses to common patient questions[7,8]. However, when applied to specific clinical contexts—such as TMD and third molar surgery—the accuracy and depth of information vary considerably[7,17]. In these areas, ChatGPT tends to provide simplified descriptions that omit critical diagnostic distinctions or therapeutic nuances. Importantly, AI-generated recommendations are inherently generalized and cannot account for individual patient characteristics, comorbidities, or psychosocial factors, underscoring the necessity of clinician-guided, individualized treatment planning in TMD management. The pronounced biopsychosocial and hormonal heterogeneity of TMD provides a clear rationale for the limited clinical applicability of generalized AI–generated outputs. Galhardo et al. reported that post-COVID-19 emotional stress significantly increased self-reported TMD pain among medical students, highlighting the psychosocial component of TMD pathophysiology[18]. Increased testosterone receptor expression in masticatory muscles with aging may contribute to sex-related differences in temporomandibular muscle dysfunction[19]. Menopausal hormonal fluctuations have been shown to correlate with the intensity of TMD-related pain, particularly during the late menopausal transition[20]. Although the model demonstrates potential as a supportive adjunct for patient education, its lack of precision and inability to reference scientific evidence underscore the need for expert oversight and evidence-based validation before clinical integration. Overall, these findings suggest that while ChatGPT can support patient education by providing accessible and generally accurate information on TMD, its outputs should be interpreted cautiously and within a clinician-guided framework.

This study has several limitations that should be acknowledged. Because the questions analyzed in this study were generated by an AI model rather than collected directly from patients, they may not fully capture the diversity, nuance, and variability of real patient perspectives and information-seeking behaviors. First, only the initial, single-turn responses from ChatGPT were analyzed; therefore, the potential variability and depth that could emerge through multi-turn or user-guided interactions were not explored. Second, the dataset was generated entirely through AI prompting, which introduces a form of authority bias, as the questions themselves were not patient-derived. Third, although five independent experts were employed to minimize subjective bias, individual professional perspectives may still have influenced scoring. Finally, the study utilized a specific version of the model (ChatGPT-4.0), and subsequent updates may alter performance characteristics. Future studies should specifically collect and analyze real patient-generated questions and directly compare them with AI-generated question sets to better assess representativeness, clinical relevance, and external validity. In addition, cross-model comparisons and longitudinal assessments may further enhance the evaluation of AI reliability in patient education.

CONCLUSION AND RECOMMENDATIONS

This study demonstrates that large language models such as Chat Generative Pre-trained Transformer (ChatGPT) possess promising potential as adjunct tools for patient education and health communication in the context of TMD. While the model generally provides accurate and understandable information,

its clinical specificity and evidential reliability remain limited. Therefore, ChatGPT and similar systems should be regarded as educational aids that complement—but do not replace—professional expertise[21,22]. AI-generated patient information should be reviewed by qualified clinicians prior to clinical use to ensure accuracy, contextual relevance, and patient safety. Future research should prioritize developing clinically validated frameworks, transparent source citation mechanisms, and patient-centered evaluation methods to ensure safe and responsible integration of AI tools into healthcare communication.

ETHICAL APPROVAL

This study did not involve human participants, patient data, or animal subjects, and was therefore exempt from institutional ethics committee review. The analysis was conducted exclusively on data generated by a publicly accessible AI model (ChatGPT). All procedures were performed in accordance with academic research integrity and transparency standards, and no identifiable or sensitive information was used.

AUTHORS' CONTRIBUTIONS

BSU: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **EEOU:** Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Sanchla AD, Shrivastav S, Kamble RH, Nerurkar SA, Toshniwal N. Altered quality of life in patients with temporomandibular joint disorders: a review. J Clin Diagn Res. 2023;17(3):ZE08-13. https://doi.org/10.7860/JCDR/2023/58474.17584
2. Warzocha J, Gadomska-Krasny J, Mrowiec J. Etiologic factors of temporomandibular disorders: a systematic review of literature containing diagnostic criteria for temporomandibular disorders (DC/TMD) and research diagnostic criteria for temporomandibular disorders (RDC/TMD) from 2018 to 2022. Healthcare (Basel). 2024;12(5):575. https://doi.org/10.3390/healthcare12050575
3. Yap AU, Lei J, Liu C, Fu KY. Characteristics of painful temporomandibular disorders and their influence on jaw functional limitation and oral health-related quality of life. J Oral Rehabil. 2024;51(9):1748-58. https://doi.org/10.1111/joor.13768
4. Fang ZY, Yang Y, Yao Y, Liu SS, Liu LK, Lu SJ, et al. Oral behaviors and anxiety are significant predictors of jaw function limitation in patients with anterior disc displacement without reduction. Arch Oral Biol. 2024;166:106033. https://doi.org/10.1016/j.archoralbio.2024.106033
5. Adams MCB, Nelson AM, Narouze S. Daring discourse: artificial intelligence in pain medicine, opportunities and challenges. Reg Anesth Pain Med. 2023;48(9):439-42. https://doi.org/10.1136/rapm-2023-104526
6. Eguia H, Sánchez-Bocanegra CL, Vinciarelli F, Alvarez-Lopez F, Saigí-Rubió F. Clinical decision support and natural language processing in medicine: systematic literature review. J Med Internet Res. 2024;26:e55315. https://doi.org/10.2196/55315
7. Aguiar Sousa R, Costa SM, Almeida Figueiredo PH, Camargos CR, Ribeiro BC, Alves E Silva MRM. Is ChatGPT a reliable source of scientific information regarding third-molar surgery? J Am Dent Assoc. 2024;155(3):227-32.e6. https://doi.org/10.1016/j.adaj.2023.11.004
8. Jeong H, Han SS, Yu Y, Kim S, Jeon KJ. How well do large language model-based chatbots perform in oral and maxillofacial radiology? Dentomaxillofac Radiol. 2024;53(6):390-5. https://doi.org/10.1093/dmfr/twae021
9. Choudhury A, Elkefi S, Tounsi A. Exploring factors influencing user perspective of ChatGPT as a technology that assists in healthcare decision making: a cross sectional survey study. PLoS One. 2024;19(3):e0296151. https://doi.org/10.1371/journal.pone.0296151
10. Johns WL, Kellish A, Farronato D, Ciccotti MG, Hammoud S. ChatGPT can offer satisfactory responses to common patient questions regarding elbow ulnar collateral ligament reconstruction. Arthrosc Sports Med Rehabil. 2024;6(2):100893. https://doi.org/10.1016/j.asmr.2024.100893
11. Özbek EA, Ertan MB, Kından P, Karaca MO, Gürsoy S, Chahla J. ChatGPT can offer at least satisfactory responses to common patient questions regarding hip arthroscopy. Arthroscopy. 2025;41(6):1806-27. https://doi.org/10.1016/j.arthro.2024.08.036
12. Li LT, Sinkler MA, Adelstein JM, Voos JE, Calcei JG. ChatGPT responses to common questions about anterior cruciate ligament reconstruction are frequently satisfactory. Arthroscopy. 2024;40(7):2058-66. https://doi.org/10.1016/j.arthro.2023.12.009
13. Mika AP, Martin JR, Engstrom SM, Polkowski GG, Wilson JM. Assessing ChatGPT responses to common patient questions regarding total hip arthroplasty. J Bone Joint Surg Am. 2023;105(19):1519-26. https://doi.org/10.2106/JBJS.23.00209
14. WordCalc. Readability calculator. 2024. [cited 2024 Dec 29]. Available from: https://www.wordcalc.com/readability/
15. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155-63. https://doi.org/10.1016/j.jcm.2016.02.012
16. Hurley ET, Crook BS, Lorentz SG, Danilkowicz RM, Lau BC, Taylor DC, et al. Evaluation high-quality of information

from ChatGPT (artificial intelligence-large language model) artificial intelligence on shoulder stabilization surgery. Arthroscopy. 2024;40(3):726-31.e6. https://doi.org/10.1016/j.arthro.2023.07.048

17. Kula B, Kula A, Bagcier F, Alyanak B. Artificial intelligence solutions for temporomandibular joint disorders: contributions and future potential of ChatGPT. Korean J Orthod. 2025;55(2):131-41. https://doi.org/10.4041/kjod24.106
18. Galhardo APM, Andrade PR, Andrade LP, Cury MAA, Mukai MK, Baracat EC, et al. Cross-sectional study of self-reported pain related to temporomandibular disorders and emotional state of medical school faculty and students: post-COVID-19 pandemic. PLoS One. 2024;19(8):e0308988. https://doi.org/10.1371/journal.pone.0308988
19. Galhardo APM, Mukai MK, Mori M, Carvalho KC, Baracat MCP, Simões MJ, et al. Influence of age and gender on sex steroid receptors in rat masticatory muscles. Sci Rep. 2019;9(1):18403. https://doi.org/10.1038/s41598-019-54774-y
20. Galhardo APM, Mukai MK, Baracat MCP, Fonseca AM, Roa CL, Sorpreso ICE, et al. Does temporomandibular disorder correlate with menopausal symptoms? Menopause. 2022;29(6):728-33. https://doi.org/10.1097/GME.0000000000001962
21. Varady NH, Lu AZ, Mazzucco M, Dines JS, Altchek DW, Williams RJ, et al. Understanding how ChatGPT may become a clinical administrative tool through an investigation on the ability to answer common patient questions concerning ulnar collateral ligament injuries. Orthop J Sports Med. 2024;12(7):23259671241257516. https://doi.org/10.1177/23259671241257516
22. Keskin T, Ünüvar BS, Ergan M, Başkurt Z, Başkurt F. Can ChatGPT provide reliable answers to frequently asked questions about physical activity? J Public Health. 2025. https://doi.org/10.1007/s10389-025-02585-z

REVIEW ARTICLE

https://doi.org/10.1590/1806-9282.20251636

Artificial intelligence to meet the psychosocial needs of cancer patients: a systematic review of randomized controlled trials

Nevra Didem Kılınç[1,2]*, Fahriye Oflaz[1]

INTRODUCTION

Cancer remains one of the leading causes of morbidity and mortality worldwide, and beyond its physical burden, the disease imposes considerable psychosocial challenges[1]. Patients frequently experience anxiety, depression, distress, and impaired quality of life throughout the diagnostic and treatment trajectory[2]. Psychosocial support is therefore recognized as an essential component of comprehensive cancer care[3]. However, access to such support is often limited due to shortages of mental health professionals, financial and institutional barriers, and the stigma still associated with seeking psychological help[4].

In recent years, artificial intelligence (AI) has emerged as a promising tool to expand the reach of supportive care[5]. AI-powered interventions such as chatbots, mobile applications, and digital therapeutics can provide personalized, scalable, and cost-effective psychosocial support[6,7]. These tools represent a technological spectrum ranging from simple rule-based algorithms to advanced generative models[5]. Crucially, the literature frequently conflates basic automation with genuine machine learning (ML), creating conceptual heterogeneity that complicates interpretation.

Although individual trials suggest potential benefits, no systematic synthesis of randomized evidence has yet clarified the effectiveness. These gaps highlight the importance of systematically reviewing and evaluating AI-based interventions to determine their effectiveness in improving psychosocial well-being among cancer patients.

METHODS

Protocol and registration

This systematic review followed Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) guidelines and was registered in PROSPERO (CRD420251134621). Literature search, study selection, data extraction, and quality assessment were performed independently by two researchers (NDK and FO), with disagreements resolved through consensus meetings.

Eligibility criteria

Studies were included if they met the PICOS framework: (P) patients diagnosed with cancer (any age, type, or stage); (I) AI-based interventions, defined operationally as tools utilizing ML (natural language processing, predictive modeling, and personalization based on behavioral data) or generative AI (large language models) to deliver psychosocial support; (C) usual care, standard psychosocial support, non-AI interventions; (O) psychosocial outcomes; (S) randomized controlled trials (RCTs) published in English. Exclusion criteria were unclear methodology, unavailable full text, observational designs, and simple rule-based algorithms or static digital health tools classified as non-AI digital interventions.

Search strategy

A literature search was conducted in PubMed, Cochrane Library, Scopus, Web of Science, and CINAHL. Keywords included "cancer," "oncology," "artificial intelligence," "digital health," "mental health," "psychological intervention," "psychosocial intervention," "mental health support," "quality of life," "anxiety," and "depression."

Data extraction

A standardized data extraction form was developed to collect key study characteristics. Extracted information included author details, publication year, country of study, study design, sample size, participant characteristics, intervention type and duration, comparison group, data collection period, and reported

[1]Koç University, Institute of Health Sciences – Istanbul, Türkiye.
[2]Istanbul Arel University, Department of Nursing – Istanbul, Türkiye.
*Corresponding author: nozluk23@ku.edu.tr
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on November 26, 2025. Accepted on January 30, 2026.
Scientific Editor: José Maria Soares Júnior

psychosocial outcomes. The data extraction process was conducted independently by two researchers (NDK and FO).

Quality assessment

The methodological quality of the included RCTs was assessed independently by two researchers using the Cochrane Risk of Bias 2 (RoB 2)[8] tool. Studies were evaluated across five domains: randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result, with overall judgments assigned accordingly. Disagreements between researchers were resolved by discussion.

RESULTS

Search results

A search of five databases revealed a total of 996 relevant articles. After removing duplicates, 143 studies were selected for full-text analysis, and only seven[9-15] met the selection criteria for inclusion in this review (Table 1). The detailed PRISMA flowchart is shown in Figure 1.

Quality assessment results

The risk of bias assessment indicated that most included trials were judged to be at low risk across all domains (Figure 2). The majority were rated overall as low risk of bias, while two trials were assessed as having some concerns due to issues related to baseline imbalances, missing outcome data, or attrition linked to higher baseline distress.

AI-based intervention results

AI-based methods in cancer care aim to enhance education, symptom management, and psychosocial support. Rule-based chatbots evaluated by Lee et al. and Shaban et al. provided information during radiotherapy or breast cancer management and reduced anxiety in some subgroups. A more advanced ChatGPT-4.0-based counseling intervention[9] significantly

Table 1. Characteristics of included studies.

Study, country	AI intervention type	Population intervention/control	Target outcome	Duration	Measurement tools
Akdogan et al.[9] Türkiye	ChatGPT-based digital counseling	Chemotherapy-naïve cancer patients 75/75	Anxiety, depression	Duration not specified; 3 time points	HADS
Jiang et al.[10], China	AI-driven mobile app (WeChat Mini Program)	Young breast cancer survivors 60/55	Psychological symptoms, self-efficacy, and social support	3-month intervention	MSAS-SF, CBI-B, SSRS, FACT-B
Kamdar et al.[11], USA	ePAL AI pain app	Metastatic solid-organ cancer patients 56/56	Pain severity, attitudes	8-week intervention	BPI, BQ-II, GAD-7, FACT-G, PHQ-8
Lee et al.[12], South Korea	Chatbot+video	Breast cancer patients 73/72	Anxiety	Duration not specified; 3 time points	APAIS, STAI, LASA
Schmitz et al.[13], Germany	Nurse AMIE virtual assistant (Alexa/Echo Show)	Metastatic breast cancer 21/21	Distress, quality of life, fatigue	6-month intervention	SF-36 Quality of Life Survey, BPI, SF-36 Pain Subscale, Sleep Disturbance Scale, SF-36 Vitality Subscale, Distress Thermometer
Shaban et al.[14], Egypt	AI-powered chatbot	Breast cancer patients 61/61	Knowledge, empowerment, attitude	Duration not specified; 2 time points	Knowledge Assessment, AI Support and Empowerment, Likert Scale for Attitude Towards AI
Springer et al.[15], Germany	Mika app-based digital therapeutic	All tumor types 99/119	Distress	12-week intervention	NCCN, HADS, FACIT-F, CGI-I

APAIS: Amsterdam Pre-operative Anxiety and Information Scale; BPI: Brief Pain Inventory Short Form Composite Severity Score; BQ-II: Barriers Questionnaire-II; CBI-B: Cancer Behavior Inventory-Brief Version; CGI-I: Clinical Global Impression-Improvement Scale; FACIT-F: Functional Assessment of Chronic Illness Therapy-Fatigue; FACT-B: Functional Assessment of Cancer Therapy-Breast; FACT-G: Functional Assessment of Cancer Therapy; GAD-7: Generalized Anxiety Disorder; HADS: Hospital Anxiety and Depression Scale; LASA: Linear Analog Scale Assessment; MSAS-SF: Memorial Symptom Assessment Scale-Short Form; NCCN: National Comprehensive Cancer Network Distress Thermometer; PHQ-8: Patient Health Questionnaire-8; SSRS: Social Support Rating Scale; STAI: Korean Short Form of the State-Trait Anxiety Inventory.; SF-36: The 36-Item Short Form Health Survey.

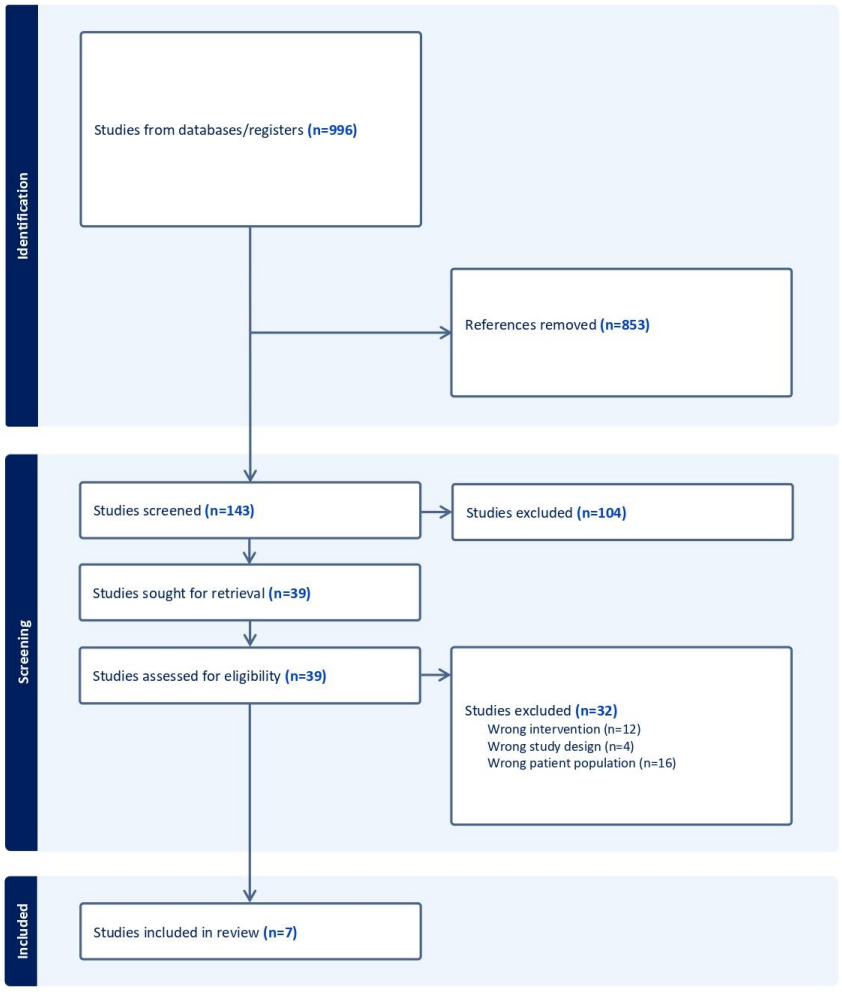


Figure 1. PRISMA flowchart.

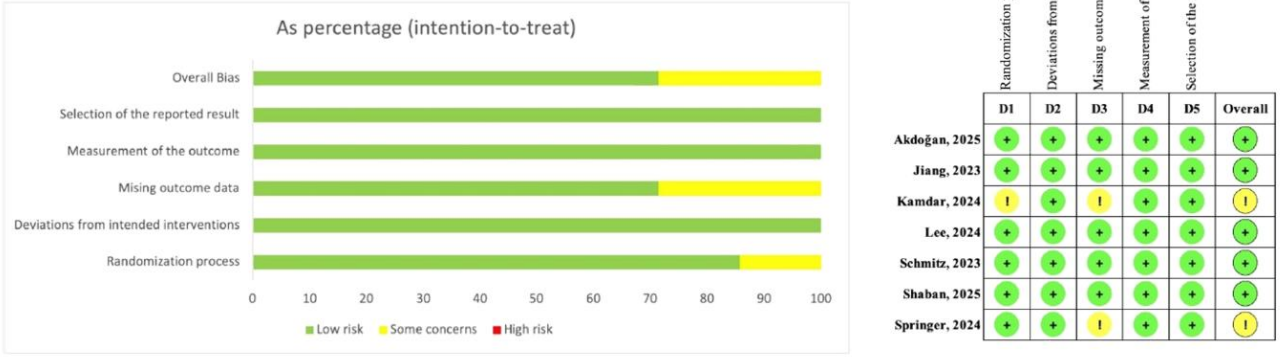


Figure 2. Risk of bias assessment (Risk of Bias 2).

reduced anxiety and depression before chemotherapy, emphasizing the need for expert validation. In a study by Jiang et al., it has been reported that "AI-TA" mobile app delivered personalized content, symptom tracking, and social support, improving psychological symptoms and self-efficacy. In an advanced cancer study by Kamdar et al., the ePAL app used AI to triage patient-reported outcomes, reducing pain scores and hospitalizations. Schmitz et al. tested the Nurse AMIE voice assistant, which showed high acceptability but no significant symptom effects. A study by Springer et al. reported that Mika app tailored support to cancer type, treatment stage, and user behavior, leading to reductions in distress, depression, anxiety, and fatigue. Only Jiang et al. explicitly applied a theoretical framework (person-centered care), while two studies incorporated therapeutic models like cognitive behavioral therapy and mindfulness-based stress reduction directly into the intervention content. Collectively, these approaches demonstrate AI's potential to deliver accessible, personalized psychosocial care, though expert oversight remains essential, especially for generative AI.

Psychosocial outcomes

Lee et al. tested a chatbot and video education during radiotherapy in breast cancer patients; while no overall effect was observed, younger patients (age <50 years) showed reduced anxiety. Akdogan et al. demonstrated that a ChatGPT-based counseling tool significantly reduced pre-chemotherapy anxiety and depression. Jiang et al. reported that the AI-TA mobile app improved psychological symptoms, self-efficacy, and social support in young breast cancer survivors. Kamdar et al. found that the ePAL app lowered pain scores and reduced pain-related hospitalizations in advanced cancer. Schmitz et al. showed that the Nurse AMIE voice assistant achieved high acceptability and satisfaction, though without significant effects on distress, symptoms, or quality of life. Shaban et al. observed that an AI chatbot significantly increased knowledge, empowerment, and positive attitudes toward AI in breast cancer patients. Springer et al. demonstrated that the Mika therapeutic app reduced distress, depression, anxiety, and fatigue in cancer patients.

DISCUSSION

This systematic review represents one of the first comprehensive syntheses of RCTs examining the effectiveness of AI-based interventions in addressing the psychosocial needs of cancer patients. The findings indicate that such digital tools show particularly promising effects in reducing anxiety and distress, while the evidence regarding depression and overall quality of life remains more limited and heterogeneous. A recent systematic review and meta-analysis demonstrated that digital psychosocial interventions (non-AI specific) significantly improved quality of life and produced moderate reductions in anxiety and depression among cancer patients, providing preliminary evidence for their integration into supportive care[16]. The trials also consistently reported high levels of acceptability, feasibility, and patient satisfaction, with interventions demonstrating the capacity to enhance knowledge, self-efficacy, and patient engagement.

International cancer guidelines, including the NCCN Distress Management Guidelines[17], ESMO Supportive Care Recommendations[18], and the ASCO Survivorship Guidelines[19], emphasize psychosocial care as an integral component of comprehensive oncology. These guidelines advocate for routine distress screening and timely psychosocial interventions for all patients. However, the translation of these recommendations into routine clinical practice remains limited due to barriers such as shortages of mental health professionals, institutional and financial constraints, and the persistent stigma associated with seeking psychological support[20]. The findings of this review suggest that AI-based interventions offer a tangible mechanism to operationalize these recommendations. For instance, the automated triage algorithms evaluated in this review[11] directly address the guideline requirement for "timely intervention" by identifying high-risk symptoms in real time, effectively bypassing delays associated with traditional referral pathways. Similarly, the digital therapeutic approaches offering on-demand coping strategies[15] provide a scalable solution to the "routine support" advocated by ASCO, particularly for patients facing access barriers or stigma. Within this context, AI-based interventions, ranging from chatbots and mobile applications to digital therapeutics, offer accessible, scalable, and personalized solutions that could help bridge the gap between guideline recommendations and real-world care delivery[21].

The findings suggest that AI interventions serve well as flexible adjuncts to conventional care. However, they introduce ethical risks including algorithmic bias[22], generative AI "hallucinations"[9], and data privacy concerns[23]. Furthermore, technological disparities may exclude older or lower-socioeconomic patients[24]. Consequently, large-scale, longitudinal studies are essential to establish the sustained effectiveness and safe integration of these tools into oncology practice.

CONCLUSION

AI-based interventions show promise for alleviating anxiety and distress, though efficacy varies. Their scalability offers a

compelling opportunity to bridge service gaps and empower patients. However, integration faces significant challenges, including algorithmic bias, data privacy concerns, and digital literacy barriers. Consequently, AI should be developed as a scalable adjunct to, rather than a replacement for, professional human support. Future large-scale, longitudinal research must prioritize evaluating safety and ethical frameworks alongside clinical outcomes to ensure AI becomes a safe and integral component of comprehensive oncology care.

Limitations

Several limitations define this review. The small number of available RCTs, combined with significant heterogeneity in intervention types and outcome measures, restricted generalizability and precluded quantitative synthesis. Most interventions lacked a theoretical framework and assessed psychosocial outcomes only as secondary endpoints. Data on long-term sustainability remain scarce, with few studies providing extended follow-up. Finally, the restriction to English-language publications and major biomedical databases, excluding specialized technology repositories due to access constraints, may have introduced selection bias.

AUTHORS' CONTRIBUTIONS

NDK: Data curation, Formal Analysis, Writing – original draft.
FO: Data curation, Formal Analysis, Supervision, Writing – original draft, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Wang Y, Feng W. Cancer-related psychosocial challenges. Gen Psychiatr. 2022;35(5):e100871. https://doi.org/10.1136/gpsych-2022-100871
2. Grassi L, Caruso R, Riba MB, Lloyd-Williams M, Kissane D, Rodin G, et al. Anxiety and depression in adult cancer patients: ESMO Clinical Practice Guideline. ESMO Open. 2023;8(2):101155. https://doi.org/10.1016/j.esmoop.2023.101155
3. Taylor TR, Adams-Campbell LL, Hicks J, Barrow J, Makambi KH, Warren J, et al. The relationship between psychosocial support service utilization, coping and post-traumatic growth among breast cancer survivors. J Psychosoc Oncol. 2025;43(6):823-40. https://doi.org/10.1080/07347332.2025.2486229
4. Bergerot C, Jacobsen PB, Rosa WE, Lam WWT, Dunn J, Fernández-González L, et al. Global unmet psychosocial needs in cancer care: health policy. EClinicalMedicine. 2024;78:102942. https://doi.org/10.1016/j.eclinm.2024.102942
5. Bekbolatova M, Mayer J, Ong CW, Toma M. Transformative potential of AI in healthcare: definitions, applications, and navigating the ethical landscape and public perspectives. Healthcare (Basel). 2024;12(2):125. https://doi.org/10.3390/healthcare12020125
6. Olawade DB, Wada OZ, Odetayo A, David-Olawade AC, Asaolu F, Eberhardt J. Enhancing mental health with artificial intelligence: current trends and future prospects. J Med Surg Public Health. 2024;3:100099. https://doi.org/10.1016/j.jomsp.2024.100099
7. Sezgin MG, Bektas H. The effectiveness of patient-centered digital empowerment programs in hematological cancer care: a systematic review and meta-analysis of randomized controlled trials. Worldviews Evid Based Nurs. 2025;22(4):e70064. https://doi.org/10.1111/wvn.70064
8. Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane handbook for systematic reviews of interventions. Version 6.5. Cochrane; 2019. Available from: https://training.cochrane.org/handbook
9. Akdogan O, Uyar GC, Yesilbas E, Baskurt K, Malkoc NA, Ozdemir N, et al. Effect of a ChatGPT-based digital counseling intervention on anxiety and depression in patients with cancer: a prospective, randomized trial. Eur J Cancer. 2025;221:115408. https://doi.org/10.1016/j.ejca.2025.115408
10. Jiang L, Xu J, Wu Y, Liu Y, Wang X, Hu Y. Effects of the "AI-TA" mobile app with intelligent design on psychological and related symptoms of young survivors of breast cancer: randomized controlled trial. JMIR Mhealth Uhealth. 2024;12:e50783. https://doi.org/10.2196/50783
11. Kamdar M, Jethwani K, Centi AJ, Agboola S, Fischer N, Traeger L, et al. A digital therapeutic application (ePAL) to manage pain in patients with advanced cancer: a randomized controlled trial. J Pain Symptom Manage. 2024;68(3):261-71. https://doi.org/10.1016/j.jpainsymman.2024.05.033
12. Lee J, Byun HK, Kim YT, Shin J, Kim YB. A study on breast cancer patient care using chatbot and video education for radiation therapy: a randomized controlled trial. Int J Radiat Oncol Biol Phys. 2025;122(1):84-92. https://doi.org/10.1016/j.ijrobp.2024.12.012
13. Schmitz KH, Kanski B, Gordon B, Caru M, Vasakar M, Truica CI, et al. Technology-based supportive care for metastatic breast cancer patients. Support Care Cancer. 2023;31(7):401. https://doi.org/10.1007/s00520-023-07884-3
14. Shaban M, Osman YM, Mohamed NA, Shaban MM. Empowering breast cancer clients through AI chatbots: transforming knowledge and attitudes for enhanced nursing care. BMC Nurs. 2025;24(1):994. https://doi.org/10.1186/s12912-025-03585-w
15. Springer F, Maier A, Friedrich M, Raue JS, Finke G, Lordick F, et al. Digital therapeutic (Mika) targeting distress in patients with cancer: results from a nationwide waitlist randomized controlled trial. J Med Internet Res. 2024;26:e51949. https://doi.org/10.2196/51949
16. Zhang Y, Flannery M, Zhang Z, Underhill-Blazey M, Bobry M, Leblanc N, et al. Digital health psychosocial intervention in adult patients with cancer and their families: systematic review and meta-analysis. JMIR Cancer. 2024;10:e46116. https://doi.org/10.2196/46116

17. Riba MB, Donovan KA, Ahmed K, Andersen B, Braun I, Breitbart WS, et al. NCCN Guidelines® insights: distress management, version 2.2023. J Natl Compr Canc Netw. 2023;21(5):450-7. https://doi.org/10.6004/jnccn.2023.0026

18. Cufer T, Kosty MP, Curriculum Development Subgroup—ESMO/ASCO Global Curriculum Working Group. ESMO/ASCO recommendations for a global curriculum in medical oncology edition 2023. JCO Glob Oncol. 2023;9:e2300277. https://doi.org/10.1200/GO.23.00277

19. Andersen BL, Lacchetti C, Ashing K, Berek JS, Berman BS, Bolte S, et al. Management of anxiety and depression in adult survivors of cancer: ASCO guideline update. J Clin Oncol. 2023;41(18):3426-53. https://doi.org/10.1200/JCO.23.00293

20. Coppini V, Ferraris G, Monzani D, Grasso R, Pravettoni G. Disparities and barriers in the assessment of psychological distress, access to and use of psycho-oncological support in Europe: current perspectives. Front Psychol. 2023;14:1252843. https://doi.org/10.3389/fpsyg.2023.1252843

21. Isa AK. Exploring digital therapeutics for mental health: AI-driven innovations in personalized treatment approaches. World J Adv Res Rev. 2024;24(3):2733-49. https://doi.org/10.30574/wjarr.2024.24.3.3997

22. Kelkar AH, Hantel A, Koranteng E, Cutler CS, Hammer MJ, Abel GA. Digital health to patient-facing artificial intelligence: ethical implications and threats to dignity for patients with cancer. JCO Oncol Pract. 2024;20(3):314-7. https://doi.org/10.1200/OP.23.00412

23. Hanna MG, Pantanowitz L, Jackson B, Palmer O, Visweswaran S, Pantanowitz J, et al. Ethical and bias considerations in artificial intelligence/machine learning. Mod Pathol. 2025;38(3):100686. https://doi.org/10.1016/j.modpat.2024.100686

24. Freitas J, Agarwal S, Schmitt B, Haslam N. Psychological factors underlying attitudes toward AI tools. Nat Hum Behav. 2023;7(11):1845-54. https://doi.org/10.1038/s41562-023-01734-2