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

RAMB

Journal of The Brazilian Medical Association

Volume 72, Number 10
October, 2026

RAMB
Journal of The Brazilian Medical Association

Volume 72, Number 10, October, 2026

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

SECTIONS

EDITORIAL

e7210EDIT Melatonin in ovarian tissue preservation: from molecular mechanisms to clinical translation

LETTERS TO THE EDITOR

e20250785 Thromboendarterectomy in unilateral chronic thromboembolism in order to improve hemodynamic and functional parameters

e20260916 The thyroid-ovary axis: beyond coincidental endocrine similarities

e20260972 ENAMED and Revalida: regulatory alignment is not psychometric equivalence

POINT OF VIEW

e20252240 Exegesis on practice-changing evidence for low-risk differentiated thyroid carcinoma in thyroidology: ab ovo usque ad mala?

SHORT COMMUNICATIONS

e20260994 Dermatological conditions in endurance athletes: prevalence and associated factors in a cross-sectional online survey

ARTICLES

ORIGINAL ARTICLES

e20251923 Do hospitalized children with obesity lose more weight than eutrophic ones?

e20260425 Associations of age and sexual abstinence duration with semen parameters in normozoospermic men

e20260486 Female sexual dysfunction in anti-synthetase syndrome

e20260531 Intrapartum rectovaginal distance as a predictor of obstetric anal sphincter injury: a prospective cohort study

e20260580 Moral distress arising from professionalism dilemmas among medical interns

e20260593 Prognostic value of the blood urea nitrogen-to-albumin ratio for 28-day mortality in septic intensive care unit patients: a large-scale retrospective cohort study using the Medical Information Mart for Intensive Care IV database

e20260627 Computed tomography of the occipital bone for sex estimation in Turks

e20260628 Systematic analysis of the scientific quality, reliability, and educational value of Vaginal Natural Orifice Transluminal Endoscopic Surgery videos on YouTube

e20260634 Evaluation of the ranawat triangle method for determining hip rotation center in total hip arthroplasty among Turkish patients

e20260659 Gastrin-releasing peptide, neuromedin B, and symptom severity in adult attention-deficit hyperactivity disorder

e20260669 The effect of micronutrient supplementation on women's sexual desire: a systematic review and meta-analysis study

e20260806 Partner hygiene–related behaviors and vaginitis subtypes: an exploratory observational study

e20260869 Comparative evaluation of ChatGPT and Gemini feedback in scenario-based clinical reasoning among medical interns

e20260876 Subacute granulomatous thyroiditis: histopathological spectrum, perineural inflammation, and coexisting lesions

e20260877 Potentially avoidable computed tomography after positive ultrasonography in appendicitis

e20260878 Magnetic resonance imaging radiomics for predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer

e20260890 A modified Rapid Emergency Medicine Score incorporating C-reactive protein for mortality prediction in septic shock

e20260891 Propofol versus remifentanil sedation for elective gastroscopy: a randomized controlled trial

e20260904 Sodium glucose cotransporter 2 inhibitors may mask malignancy-associated fluorodeoxyglucose uptake in the thyroid on positron emission tomography/computed tomography

e20260914 Opportunistic computed tomography-based vertebral attenuation for osteoporosis screening

e20260920 Digital conscience in midwifery students: a qualitative study

e20260923 Hemorrhoidectomy vs. non-Doppler-guided hemorrhoidal artery ligation

e20260932 Association between keratoconus and arterial stiffness: a cross-sectional study

ERRATUM

E20260070-ER ERRATUM: Association between lower limb strength, anxiety, and depression symptoms in older adults

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 Deláscio 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/10.1590/1806-9282.7210EDIT

Melatonin in ovarian tissue preservation: from molecular mechanisms to clinical translation

Karla Krislane Alves Costa Monteiro[1], Edmund Chada Baracat[1], José Maria Soares Júnior[1]*

Ovarian tissue cryopreservation and transplantation represent the only fertility preservation strategy available for prepubertal cancer patients and women requiring urgent gonadotoxic therapy. However, cryoinjury, ischemia-reperfusion damage, and oxidative stress significantly compromise follicular viability and graft longevity. Melatonin, a multifunctional indoleamine with potent antioxidant properties, has emerged as a promising adjunct to mitigate these challenges.

In a recently published systematic review, Monteiro et al.[1] comprehensively evaluated the antioxidant actions of melatonin in animal models of ovarian tissue preservation. The analysis revealed that melatonin consistently reduces oxidative stress markers and enhances the expression and activity of key antioxidant enzymes across multiple experimental models. The systematic synthesis demonstrated that melatonin's efficacy is dose-dependent and varies according to administration route, with intraperitoneal and subcutaneous routes showing particularly robust effects. Importantly, the review identified that melatonin's antioxidant actions extend beyond direct radical scavenging to include the modulation of nuclear factor erythroid 2-related factor 2 (Nrf2)-mediated signaling pathways and the preservation of mitochondrial function[1].

Oxidative stress represents a central pathophysiological mechanism in ovarian tissue injury during cryopreservation and transplantation. The generation of reactive oxygen species (ROS) during freeze-thaw cycles and subsequent reperfusion triggers lipid peroxidation, mitochondrial dysfunction, and DNA damage in granulosa cells and oocytes[2]. Melatonin exerts potent antioxidative effects through multiple pathways, significantly reducing ROS concentrations and improving follicular morphological integrity[3].

A pivotal contribution from our group demonstrated that melatonin increases superoxide dismutase 2 (SOD2) levels and improves rat ovarian graft function after transplantation[4]. In this study, female rats underwent bilateral ovariectomy followed by heterotopic transplantation of frozen-thawed ovarian tissue. Melatonin was administered intraperitoneally at 10 mg/kg for 21 days post-transplantation. The results were striking: melatonin-treated animals exhibited significantly higher SOD2 immunoexpression in theca interna cells and interstitial cells, accompanied by reduced follicle-stimulating hormone (FSH) levels and favorable FSH/anti-Müllerian hormone ratios. Histological analysis revealed that melatonin-treated grafts contained more mature follicles, showed enhanced type I collagen deposition, and demonstrated improved angiogenesis compared to controls[4]. These findings are clinically significant because SOD2, the mitochondrial isoform of superoxide dismutase, serves as the primary defense against mitochondrial superoxide radicals. The preservation of SOD2 activity is crucial for maintaining mitochondrial homeostasis and preventing the oxidative damage that compromises granulosa cell function and oocyte quality.

Expanding upon these findings, our subsequent investigation examined the effects of melatonin on glutathione peroxidase (GPx) activity and overall antioxidant defense in cryopreserved ovarian transplants55. Melatonin treatment significantly increased GPx activity in grafted ovarian tissues, accompanied by reduced malondialdehyde levels, indicating decreased lipid peroxidation. Furthermore, melatonin upregulated the expression of GPx4, the selenoenzyme isoform critical for protecting cells against ferroptosis—an iron-dependent, oxidative stress-driven form of cell death[5]. The enhancement of GPx activity represents a complementary mechanism to SOD2 upregulation, as both enzymes work sequentially to neutralize ROS: SOD2 converts superoxide radicals to hydrogen peroxide, which is subsequently detoxified by GPx to water. This coordinated antioxidant defense is essential for maintaining the redox balance in grafted ovarian

[1]Laboratório de Investigação em Ginecologia Molecular e Estrutural da Disciplina de Ginecologia, Departamento de Obstetrícia e Ginecologia, Faculdade de Medicina da Universidade de São Paulo – São Paulo (SP), Brazil.
*Corresponding author: jsoares415@hotmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on July 04, 2026. Accepted on July 04, 2026.
Scientific Editor: Roseli Nomura

tissues, where ischemia-reperfusion injury creates a burst of oxidative stress.

A particularly innovative contribution from our group explored the use of scaffold-based melatonin delivery systems for frozen-thawed ovarian autografts[6]. This study represents a significant translational advance, moving beyond systemic melatonin administration toward localized, sustained delivery directly to the graft site. Using Gelfoam® as a biodegradable scaffold loaded with melatonin, the researchers demonstrated that local delivery effectively reduced oxidative stress and apoptosis while promoting angiogenesis in transplanted ovarian tissues. The scaffold-based approach offers several advantages: it ensures high local melatonin concentrations at the graft site, minimizes systemic exposure, provides sustained release over the critical early post-transplantation period, and enables non-invasive application. Histological and immunohistochemical analyses revealed that scaffold-delivered melatonin preserved follicular morphology, increased vascular endothelial cell density, and reduced inflammatory cytokine levels compared to controls. Endocrine assessment showed accelerated recovery of estrous cycles, with treated animals achieving regular cyclicity earlier than untreated animals. The scaffold-based approach also demonstrated superior preservation of primordial and primary follicles, which represent the ovarian reserve and are crucial for long-term graft function[6].

The collective evidence from our research group and others supports a multifaceted mechanism of melatonin action in ovarian tissue preservation. Melatonin preserves mitochondrial morphology and function, as evidenced by increased ATP levels and succinate dehydrogenase activity[2]. Beyond SOD2 the Nrf2/antioxidant response element signaling pathway, upregulating downstream antioxidant enzymes including heme oxygenase-1 and catalase[7]. The upregulation of GPx4 by melatonin is particularly significant for inhibiting ferroptosis, a recently recognized form of cell death characterized by iron-dependent lipid peroxidation[5]. Melatonin also promotes angiogenesis through multiple mechanisms, including increased vascular endothelial growth factor expression and enhanced SOD2 activity in endothelial cells, which is critical for graft survival[4]. Finally, the combined antioxidant and angiogenic effects of melatonin result in reduced apoptosis in granulosa cells, theca cells, and interstitial cells[4,7,8].

Melatonin represents a compelling therapeutic adjunct in ovarian tissue preservation, acting through multiple complementary mechanisms—direct ROS scavenging, SOD2 and GPx upregulation, Nrf2 pathway activation, ferroptosis inhibition, and promotion of angiogenesis. Our systematic review has confirmed the robust antioxidant actions of melatonin across multiple animal models[1], while our experimental studies have established its ability to increase SOD2 levels and improve graft function44, enhance GPx activity and antioxidant defense55, and achieve effective graft protection through scaffold-based delivery66. As ovarian tissue cryopreservation continues to expand in clinical practice, incorporating melatonin into preservation protocols—whether through systemic administration or innovative scaffold-based delivery—offers a safe, cost-effective strategy to maximize follicular survival and graft longevity. Future research should prioritize translational studies and clinical trials to establish evidence-based guidelines, ultimately improving fertility outcomes for young women facing gonadotoxic therapies.

AUTHORS' CONTRIBUTIONS

KKACM: Conceptualization; Data curation; Formal Analysis; Investigation; Methodology; Project administration; Resources; Validation; Visualization; Writing – original draft; Writing – review & editing. **ECB:** Conceptualization; Funding acquisition; Project administration; Resources; Supervision; Validation; Visualization; Writing – review & editing. **JMSJ:** Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; 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. Monteiro KKAC, Shiroma ME, Damous LL, Simões MJ, Simões RDS, Cipolla-Neto J, et al. Antioxidant actions of melatonin: a systematic review of animal studies. Antioxidants (Basel). 2024;13(4):439. https://doi.org/10.3390/antiox13040439
2. Sun TC, Liu XC, Yang SH, Song LL, Zhou SJ, Deng SL, et al. Melatonin inhibits oxidative stress and apoptosis in cryopreserved ovarian tissues via Nrf2/HO-1 signaling pathway. Front Mol Biosci. 2020;7:163. https://doi.org/10.3389/fmolb.2020.00163
3. Alkali IM, Colombo M, Luvoni GC. Melatonin reduces oxidative stress and improves follicular morphology in feline (Felis catus) vitrified ovarian tissue. Theriogenology. 2024;224:58-67. https://doi.org/10.1016/j.theriogenology.2024.05.011
4. Monteiro KKAC, Damous LL, Shiroma ME, Termini L, Cipolla-Neto J, Simões RDS, et al. Melatonin increases superoxide dismutase 2 (SOD2) levels and improves rat ovarian graft

function after transplantation. J Ovarian Res. 2024;17(1):204. https://doi.org/10.1186/s13048-024-01512-2

5. Monteiro KKAC, Damous LL, Shiroma ME, Turri JAO, Simões RDS, Simões MJ, et al. Melatonin enhances glutathione peroxidase activity and improves antioxidant defense in cryopreserved ovarian transplants: a rat model study. Antioxidants (Basel). 2026;15(5):551. https://doi.org/10.3390/antiox15050551
6. Damous LL, Alves MS, Justino LG, Son ACH, Botelho NM, Shiroma ME, et al. Frozen-Thawed ovarian autografts treated with scaffold-based melatonin delivery in rats. Reprod Biol Endocrinol. 2026;24(1):38. https://doi.org/10.1186/s12958-026-01540-1
7. Xie L, Wang S, Wu Y, Zhang X, Ding Y. Melatonin protects intact rat ovarian transplantation via the MT1/Nrf2/ARE pathway. Cells. 2025;14(20):1588. https://doi.org/10.3390/cells14201588
8. Shiroma ME, Damous LL, Cotrim FP, Roa CL, Cipolla-Neto J, Reiter RJ, et al. Pretreatment with melatonin improves ovarian tissue cryopreservation for transplantation. Reprod Biol Endocrinol. 2021;19(1):17. https://doi.org/10.1186/s12958-021-00705-4

https://doi.org/10.1590/1806-9282.20250785

Thromboendarterectomy in unilateral chronic thromboembolism in order to improve hemodynamic and functional parameters

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

Dear Editor,

We have read the research article "Unilateral Chronic Thromboembolic Pulmonary Disease: Do Patients Benefit From Thromboendarterectomy with great interest." This utilitarian research seems to demand a determination that sheds light on the efficacy of the pulmonary artery bypass graft in order to treat this unusual pathologic phenomenon, unilateral chronic thromboembolism pulmonary disease (CTEPD), though certain restrictions are ineludible, which affects the soundness of the outcomes in the issue of thromboembolism[1-3]. Rashidi et al.[1] reported this multicentre study, involving the cases with unilateral CTEPD from three referral centers, evaluating their demographic information, 6-min walk test distance (6MWD), New York Heart Association (NYHA) functional class, and hemodynamics. The authors emphasized a significant progression in the mean pulmonary artery pressure and pulmonary vascular resistance. They also remarked on a substantial promotion in the NYHA functional class and 6MWD. Rashidi et al.[1] re-evaluated the cases for at least 3 months after discharge, and pre- and postoperative 6MWD data were performed. The authors analyzed the pathology in the studied population using the 6MWT 3 months after surgery, and the mortality was appraised in the hospital and 1 year after surgery. The study has several notable strengths. Firstly, it focuses on a very rare condition, unilateral CTEPD, thereby contributing valuable data to a scarcely populated area of research. The multicenter design enhances the generalizability of the findings. Furthermore, the inclusion of patients with CTEPD without pulmonary hypertension (PH), who showed symptomatic improvement despite normal resting hemodynamics, is a significant contribution. The authors also rightly emphasize the importance of differentiating unilateral pulmonary artery involvement from other etiologies.

However, the study also has some flaws and limitations that warrant consideration. Initially, the sample size, while significant for such a rare condition, remains relatively small, which may limit the number and affect the statistical power of subgroup analyses, such as the comparison between lobar and segmental involvement. Secondly, the lack of postoperative right heart catheterization data at 3 months for all the cases, as well as the use of different cardiac output measurement methods across the centers, could introduce variability in the assessment of haemodynamic outcomes. Furthermore, the study does not delve into the etiology of unilateral involvement. While the authors mention previous reports suggesting a link between inherited thrombophilia and unilateral CTEPH in young women, the current study only assessed for antiphospholipid syndrome[1]. Further investigation into other potential risk factors, such as other thrombophilias and the role of estrogen-based medications in women, would be beneficial. Nevertheless, would the utilization of long-term follow-up beyond a 1-year period for all the studied participants alter the outcome(s) of this study? Furthermore, would harnessing larger sampling prior to the procedure transmute the desinence(s) of this work?

In conclusion, this study provides substantial evidence supporting the safety and efficacy of pulmonary artery bypass graft in improving symptoms and haemodynamic parameters in carefully selected patients with unilateral CTEPD, including those without PH. However, the limitations of the study, particularly the small sample size, should be considered when interpreting the results[1]. Larger, prospective studies with standardized follow-up protocols are necessary to validate these findings further and to gain a better understanding of the pathophysiology and long-term outcomes of unilateral CTEPD. It is also crucial to continue to refine the diagnostic work-up to differentiate unilateral CTEPD from other conditions affecting the pulmonary

[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 23, 2025. Accepted on June 28, 2025.
Scientific Editor: José Maria Soares Júnior

arteries. This issue merits further investigation. We thank Rashidi et al.[1] for their valuable study on thromboembolism[2,3].

AUTHORS' CONTRIBUTIONS

IS: Conceptualization, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **DS:** Conceptualization, 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. Rashidi F, Yıldızeli B, Parvizi R, Taş S, Yıldızeli ŞO, Mutlu B, et al. Unilateral chronic thromboembolic pulmonary disease: do patients benefit from thromboendarterectomy? Case series from three CTEPH centres. Heart Lung Circ. 2024;33:1574-81 https://doi.org/10.1016/j.hlc.2024.06.049
2. Mazurová T, Sengul I, Toman D, Pelikán A, Sengul D, Miloslav M, et al. Endofibrosis as a causative agent of the peripheral artery disease: a report of two cases for professional cyclists. Cureus. 2022;14:e31406. https://doi.org/10.7759/cureus.31406
3. Mazurová T, Sengul I, Toman D, Pelikán A, Sengul D, Miloslav M, et al. Endofibrosis as a cause of the peripheral artery disease: a comprehensive review and proposal of two novel algorithms for diagnosis and treatment. Rev Assoc Med Bras. 2023;69:352-6. https://doi.org/10.1590/1806-9282.202201374

https://doi.org/10.1590/1806-9282.20260916

The thyroid-ovary axis: beyond coincidental endocrine similarities

Gulru Birce Sonmezoz[1]*, Evrim Cakir[2]

Dear Editor,

Although the thyroid gland and ovary are traditionally considered distinct endocrine organs, accumulating evidence suggests that they share structural, hormonal, and functional characteristics. Recognition of these similarities offers a useful perspective for understanding several clinical observations involving both thyroid and reproductive health.

A notable structural similarity between the thyroid gland and ovary is the presence of a predominantly simple cuboidal epithelium. However, the ovarian surface epithelium ranges from squamous to cuboidal, whereas the thyroid follicular epithelium ranges from squamous to columnar depending on functional activity[1,2].

Beyond this histological resemblance, both tissues demonstrate responsiveness to common endocrine signals. Human ovarian tissue, including oocytes, granulosa cells, stromal cells, and surface epithelial cells, has been shown to express thyroid-stimulating hormone (TSH) receptors, while thyroid hormone receptors are expressed in granulosa cells, cumulus cells, and oocytes[3,4]. Furthermore, thyroid hormones are detectable within follicular fluid, suggesting a direct role in ovarian physiology[5]. An experimental study has demonstrated that triiodothyronine (T3) influences gene expression associated with inflammation and estrogen receptor (ER) α expression in ovarian surface epithelial cells[6], suggesting that thyroid hormones may actively participate in ovarian cellular regulation rather than exerting solely indirect systemic effects. An association between ovarian cancer and hyperthyroidism has been reported in some studies, although the underlying mechanisms remain unclear and may involve shared hormonal signaling pathways[6,7].

On the other hand, thyroid tissue is responsive to reproductive hormones. Estrogen and progesterone modulate thyroid follicular cell activity[8,9], and both estrogen and progesterone receptors are frequently expressed in papillary thyroid carcinoma[10]. Estradiol has been shown to stimulate thyroid cell proliferation and may contribute to thyroid growth under certain physiological and pathological conditions, including papillary thyroid cancer[11,12]. Interestingly, oral contraceptive use has been associated with reduced thyroid volume and a lower prevalence of goiter[13]. These apparently divergent findings may also reflect differences in the type, dose, and route of administration of estrogenic compounds[14]. An experimental animal study has demonstrated that conjugated equine estrogens induce a hyperstimulatory effect on thyroid morphology, characterized by reduced follicular and colloid areas, increased epithelial height, and enhanced proliferative activity, while tamoxifen exerted distinct histomorphological effects, suggesting that the structural impact of estrogenic compounds on the thyroid is compound-specific[15].

While increased thyroid volume has been reported in women with polycystic ovary syndrome (PCOS), which is characterized by hormonal fluctuations, exogenous estrogen therapy is typically administered in a more stable hormonal context. This suggests that endogenous and exogenous hormonal exposures may exert distinct biological effects. In particular, continuous exogenous estrogen therapy is associated with suppression of luteinizing hormone (LH), which has been proposed to have independent effects on thyroid physiology. Therefore, differences in thyroid outcomes between these states likely reflect variations in overall endocrine feedback regulation and gonadotropin dynamics rather than estradiol exposure alone. Consistent with this interpretation, a large prospective study in women with hysterectomy alone found no association between hormone replacement therapy and thyroid cancer risk[16].

The existence of struma ovarii further illustrates the biological overlap between these organs. As a specialized ovarian teratoma predominantly composed of thyroid tissue, struma ovarii demonstrates that thyroid tissue can develop and remain functionally active within the ovary. The biochemical manifestations

[1]Taksim Training and Research Hospital, Department of Internal Medicine – Istanbul, Turkey.
[2]Taksim Training and Research Hospital, Department of Endocrinology and Metabolic Diseases – Istanbul, Turkey.
*Corresponding author: gulrubirces@gmail.com

Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on July 02, 2026. Accepted on August 04, 2026.
Scientific Editor: José Maria Soares Júnior

of hyperthyroidism associated with struma ovarii closely resemble those observed in primary hyperthyroidism[17].

Shared hormonal signaling pathways provide additional evidence for this relationship. TSH, follicle-stimulating hormone (FSH), and LH are structurally similar heterodimeric glycoproteins and consist of an identical α-subunit with structurally distinct β-subunits[18]. FSH has been shown in experimental animal models to induce thyroid follicular changes resembling those observed following TSH stimulation[19]. Experimental studies have demonstrated that LH can stimulate thyroid adenylate cyclase activity and enhance iodine uptake by thyroid tissue[20,21]. Consistent with these findings, clinical studies have identified a positive association between thyroid volume and LH levels in women of reproductive age[22]. In contrast, the use of oral contraceptives was associated with a decreased thyroid volume and a fourfold reduction in the risk of goiter[13]. The distinct effects between endogenous estradiol fluctuation and exogenous estrogen exposure may be partly attributable to differences in LH regulation.

The relationship becomes particularly evident during pregnancy through the actions of human chorionic gonadotropin (hCG), which shares a common α-subunit with TSH, FSH, and LH. It has been shown that hCG can stimulate thyroid activity in a manner similar to TSH, leading to increased TSH receptor expression, iodine uptake, adenylate cyclase activity, and thyroid hormone secretion[22]. During early pregnancy, elevated hCG levels exert a transient thyrotropic effect via the TSH receptor, leading to stimulation of thyroid growth, increased production of T4 and T3, and subsequent suppression of TSH[23]. This interaction may represent a physiological maternal adaptation that supports increased thyroid hormone availability during early pregnancy.

Another clinically relevant link between the thyroid and ovary is the association between autoimmune thyroid disease and premature ovarian insufficiency. Autoimmune thyroiditis has been identified in approximately 14–32.7% of women with premature ovarian insufficiency at the time of diagnosis[24]. However, current evidence suggests that thyroid autoantibody positivity alone is not consistently associated with diminished ovarian reserve, indicating that accompanying thyroid dysfunction, rather than thyroid autoimmunity itself, may play a more important role in reproductive impairment[25].

Clinical studies also suggest that thyroid-directed therapies may directly affect ovarian function. Radioactive iodine (RAI) treatment has been associated with menstrual irregularities and reductions in anti-Müllerian hormone (AMH) levels, indicating a potential adverse impact on ovarian reserve[26,27]. While current evidence does not establish a direct causal relationship between all thyroid interventions and impaired fertility, these findings emphasize the importance of considering reproductive outcomes when selecting treatment strategies for women of reproductive age. They also highlight the need for prospective studies evaluating the reproductive safety of emerging thyroid therapies, such as radiofrequency and microwave ablation.

In conclusion, available evidence suggests that the thyroid and ovary share more than a coincidental endocrine relationship. Structural similarities, common receptor expression patterns, overlapping hormonal signaling pathways, and shared clinical phenomena collectively support the existence of meaningful biological interactions between these organs. Although many observations remain associative and mechanistic studies are still limited, viewing thyroid–ovarian interactions within an integrated endocrine framework may provide new insights into reproductive disorders, thyroid disease, and treatment-related complications. Future research should aim to further elucidate the molecular mechanisms underlying thyroid–ovary interactions and evaluate the reproductive consequences of thyroid diseases and their treatment.

AUTHORS' CONTRIBUTIONS

GBS: Conceptualisation, Data curation, Writing – original draft. **EC:** Supervision, Writing – review & editing.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

1. Al-Suhaimi EA, Khan FA. Thyroid glands: physiology and structure. In: Al-Suhaimi EA, editor. Emerging concepts in endocrine structure and functions. Singapore: Springer; 2022. p. 133-60. https://doi.org/10.1007/978-981-16-9016-7_5
2. Auersperg N, Wong AS, Choi KC, Kang SK, Leung PC. Ovarian surface epithelium: biology, endocrinology, and pathology. Endocr Rev. 2001;22(2):255-88. https://doi.org/10.1210/edrv.22.2.0422
3. Aghajanova L, Lindeberg M, Carlsson IB, Stavreus-Evers A, Zhang P, Scott JE, et al. Receptors for thyroid-stimulating hormone and thyroid hormones in human ovarian tissue. Reprod Biomed Online. 2009;18(3):337-47. https://doi.org/10.1016/s1472-6483(10)60091-0
4. Zhang SS, Carrillo AJ, Darling DS. Expression of multiple thyroid hormone receptor mRNAs in human oocytes, cumulus cells, and granulosa cells. Mol Hum Reprod. 1997;3(7):555-62. https://doi.org/10.1093/molehr/3.7.555

5. Wakim AN, Polizotto SL, Buffo MJ, Marrero MA, Burholt DR. Thyroid hormones in human follicular fluid and thyroid hormone receptors in human granulosa cells. Fertil Steril. 1993;59(6):1187-90. https://doi.org/10.1016/s0015-0282(16)55974-3
6. Rae MT, Gubbay O, Kostogiannou A, Price D, Critchley HOD, Hillier SG, et al. Thyroid hormone signaling in human ovarian surface epithelial cells. J Clin Endocrinol Metab. 2007;92(1):322-7. https://doi.org/10.1210/jc.2006-1522
7. Wang T, Wang X, Wu J, Li X. Causal relationship between thyroid dysfunction and ovarian cancer: a two-sample Mendelian randomization study. BMC Cancer. 2024;24(1):629. https://doi.org/10.1186/s12885-024-12385-5
8. Furlanetto TW, Nguyen LQ, Jameson JL. Estradiol increases proliferation and down-regulates the sodium/iodide symporter gene in FRTL-5 cells. Endocrinology. 1999;140(12):5705-11. https://doi.org/10.1210/endo.140.12.7197
9. Bertoni AP, Brum IS, Hillebrand AC, Furlanetto TW. Progesterone upregulates gene expression in normal human thyroid follicular cells. Int J Endocrinol. 2015;2015:864852. https://doi.org/10.1155/2015/864852
10. Ahn HY, Song RY, Ahn HS, Kim HS. Expression of estrogen and progesterone receptors in papillary thyroid carcinoma in Korea. Cancer Res Treat. 2021;53(4):1204-12. https://doi.org/10.4143/crt.2020.1201
11. Derwahl M, Nicula D. Estrogen and its role in thyroid cancer. Endocr Relat Cancer. 2014;21(5):T273-83. https://doi.org/10.1530/ERC-14-0053
12. Xu FZ, Zheng LL, Chen KH, Wang R, Yi DD, Jiang CY, et al. Serum sex hormones correlate with pathological features of papillary thyroid cancer. Endocrine. 2024;84(1):148-54. https://doi.org/10.1007/s12020-023-03554-w
13. Knudsen N, Bülow I, Laurberg P, Perrild H, Ovesen L, Jørgensen T, et al. Low goitre prevalence among users of oral contraceptives in a population sample of 3712 women. Clin Endocrinol (Oxf). 2002;57(1):71-6. https://doi.org/10.1046/j.1365-2265.2002.01564.x
14. Tahboub R, Arafah BM. Sex steroids and the thyroid. Best Pract Res Clin Endocrinol Metab. 2009;23(6):769-80. https://doi.org/10.1016/j.beem.2009.06.005
15. Araujo LF, Soares JM Jr, Simões RS, Calió PL, Oliveira-Filho RM, Simões MJ, et al. Effect of conjugated equine estrogens and tamoxifen administration on thyroid gland histomorphology of the rat. Clinics (Sao Paulo). 2006;61(4):321-6. https://doi.org/10.1590/s1807-59322006000400008
16. Luo J, Hendryx M, Manson JE, Liang X, Margolis KL. Hysterectomy, oophorectomy, and risk of thyroid cancer. J Clin Endocrinol Metab. 2016;101(10):3812-9. https://doi.org/10.1210/jc.2016-2011
17. Zamani F, Abdolrazaghnejad A, Ameli F, GHashghaee S, Nassiri S, Zamani N, et al. Struma ovarii: a case report and review the literature. Int J Surg Case Rep. 2022;96:107318. https://doi.org/10.1016/j.ijscr.2022.107318
18. Wide L, Eriksson K. Molecular size and charge as dimensions to identify and characterize circulating glycoforms of human FSH, LH and TSH. Ups J Med Sci. 2017;122(4):217-23. https://doi.org/10.1080/03009734.2017.1412373
19. Csaba G. Receptorial and functional overlapping between thyrotropin (TSH) and gonadotropins (FSH, LH): phylogenetic, ontogenetic and clinical application aspects. Annals Thyroid Res. 2018;4(3):158-63. https://doi.org/10.26420/ANNALSTHYROIDRES.2018.1041
20. Carayon P, Lefort G, Nisula B. Interaction of human chorionic gonadotropin and human luteinizing hormone with human thyroid membranes. Endocrinology. 1980;106(6):1907-16. https://doi.org/10.1210/endo-106-6-1907
21. Yoshimura M, Hershman JM, Pang XP, Berg L, Pekary AE. Activation of the thyrotropin (TSH) receptor by human chorionic gonadotropin and luteinizing hormone in Chinese hamster ovary cells expressing functional human TSH receptors. J Clin Endocrinol Metab. 1993;77(4):1009-13. https://doi.org/10.1210/jcem.77.4.7691861
22. Cakir E, Sahin M, Topaloglu O, Colak NB, Karbek B, Gungunes A, et al. The relationship between LH and thyroid volume in patients with PCOS. J Ovarian Res. 2012;5(1):43. https://doi.org/10.1186/1757-2215-5-43
23. Glinoer D. The regulation of thyroid function in pregnancy: pathways of endocrine adaptation from physiology to pathology. Endocr Rev. 1997;18(3):404-33. https://doi.org/10.1210/edrv.18.3.0300
24. Szeliga A, Calik-Ksepka A, Maciejewska-Jeske M, Grymowicz M, Smolarczyk K, Kostrzak A, et al. Autoimmune diseases in patients with premature ovarian insufficiency-our current state of knowledge. Int J Mol Sci. 2021;22(5):2594. https://doi.org/10.3390/ijms22052594
25. Osuka S, Iwase A, Goto M, Takikawa S, Nakamura T, Murase T, et al. Thyroid autoantibodies do not impair the ovarian reserve in euthyroid infertile women: a cross-sectional study. Horm Metab Res. 2018;50(7):537-42. https://doi.org/10.1055/a-0637-9430
26. Sioka C, Kouraklis G, Zafirakis A, Manetou A, Dimakopoulos N. Menstrual cycle disorders after therapy with iodine-131. Fertil Steril. 2006;86(3):625-8. https://doi.org/10.1016/j.fertnstert.2006.02.081
27. Piek MW, Postma EL, Leeuwaarde R, Boer JP, Bos AME, Lok C, et al. The effect of radioactive iodine therapy on ovarian function and fertility in female thyroid cancer patients: a systematic review and meta-analysis. Thyroid. 2021;31:658. https://doi.org/10.1089/thy.2020.0356

https://doi.org/10.1590/1806-9282.20260972

ENAMED and Revalida: regulatory alignment is not psychometric equivalence

Vinícius Côgo Destefani[1], Matheus Feliciano da Costa Ferreira[1], Afrânio Côgo Destefani[2]*

Dear Editor,

For more than a decade, Brazil used separate frameworks to assess domestic medical programs and graduates of foreign medical schools. Carvalho et al.[1] documented substantial differences between the assessment blueprints. INEP Ordinance No. 478, issued in July 2025[2], introduced a Common Reference Matrix (Matriz de Referência Comum) for ENAMED and Revalida. The two item-writing committees now work from the same set of clinical domains and competencies. This is an important regulatory improvement.

Regulatory alignment, however, does not establish psychometric equivalence. A shared blueprint supports content representation, but Kane's validity framework[3] requires evidence for each intended interpretation and use of scores. Table 1 summarizes relevant differences between the two examinations.

Because the groups differ in training and clinical context, items may not function identically across examinations. Item difficulty and discrimination should therefore be examined separately. Recent work in health professions education shows how differential item functioning (DIF) analysis can identify items that behave differently across groups[4]. For example, linguistically dense stems may disadvantage candidates with less exposure to Portuguese without reflecting differences in clinical knowledge.

The Common Reference Matrix improves curricular coherence but does not establish score comparability. Evidence from medical licensure assessment also shows why item behavior should be examined across relevant groups[5]. We recommend that INEP introduce a common set of anchor items after both examinations have operated under the shared blueprint. These items could support routine DIF analyses and fairness monitoring. Item calibration and standard setting should remain specific to each examination unless evidence supports a common scale.

GENERATIVE AI STATEMENT

Generative AI tools assisted with language editing and literature searches. The authors verified the sources and take responsibility for the analysis and conclusions. AI tools were not credited as authors.

AUTHORS' CONTRIBUTIONS

VCD: Conceptualization, Formal Analysis, Writing – original draft. **MFCF:** Project administration, Writing – review & editing. **ACD:** Methodology, Validation, Writing – review & editing.

DATA AVAILABILITY STATEMENT

No new data were generated or analyzed for this correspondence.

[1]Centro Universitário Integrado – Campo Mourão (PR), Brazil.
[2]Escola Superior de Ciências da Santa Casa de Misericórdia de Vitória (EMESCAM) – Vitória (ES), Brazil.
*Corresponding author: afranio.destefani@emescam.br

Conflicts of interest: Vinícius Côgo Destefani is Director of Pedagogy and Engagement and co-founder of SPRMed (São Paulo, Brazil), and Medical Advisor at Allm Inc (Tokyo, Japan). Matheus Feliciano da Costa Ferreira is CEO of SPRMed. Afrânio Côgo Destefani serves as a contracted scientific consultant for SPRMed. The SPRMed item bank and platform are commercial products that provide formative assessment analytics. The authors declare no other competing interests. Vinícius Côgo Destefani and Afrânio Côgo Destefani have a family relationship. Funding: none.
Received on June 26, 2026. Accepted on July 06, 2026.
Scientific Editor: Roseli Nomura

Table 1. Structural asymmetries between ENAMED and Revalida.

Dimension	ENAMED (2025)	Revalida
Primary purpose	Program evaluation and domestic residency admission	Individual licensure for foreign-trained medical graduates
Target population	Domestic medical students	Foreign-trained candidates
Training context	Brazilian medical schools under national curricular guidelines	Medical schools in multiple national systems
Language and clinical context	Training within Brazilian clinical settings	Variable prior exposure to Portuguese and Brazilian clinical settings

Note: These differences are relevant to separate psychometric evaluation of the two examinations.

REFERENCES

1. Carvalho AB, Resende HL, Faria RMD, Toledo Júnior AC. Análise de conteúdo de duas avaliações externas brasileiras de cursos de medicina: Enade e Revalida. Rev Bras Educ Med. 2021;45(1):e033. https://doi.org/10.1590/1981-5271v45.1-20200365
2. Ministério da Educação (Brazil). Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Portaria nº 478, de 18 de julho de 2025. Dispõe sobre a implementação da Matriz de Referência Comum para a Avaliação da Formação Médica. Diário Oficial da União. Seção 1:40. 2025. Available from: https://www.in.gov.br/web/dou/-/portaria-n-478-de-18-de-julho-de-2025-643085112
3. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. Med Educ. 2015;49(6):560-75. https://doi.org/10.1111/medu.12678
4. Tavakol M, Stewart C, Sharpe CC. Ensuring fairness in assessment in health professions education: rapid analysis tools to detect differential item functioning across groups. Int J Med Educ. 2024;15:80-3. https://doi.org/10.5116/ijme.6694.de69
5. Rubright JD, Jodoin M, Woodward S, Barone MA. Differential item functioning analysis of United States Medical Licensing Examination Step 1 items. Acad Med. 2022;97(5):718-22. https://doi.org/10.1097/ACM.0000000000004567

https://doi.org/10.1590/1806-9282.20252240

Exegesis on practice-changing evidence for low-risk differentiated thyroid carcinoma in thyroidology: ab ovo usque ad mala?

Ilker Sengul[1,2,3]*, Demet Sengul[4]

Dear Editor,

Ad augusta per angusta! We have perused with palpable interest the illuminating comment penned by Hartl[1], bearing the title "Practice-changing evidence for low-risk differentiated thyroid cancer" in thyroidology[1-6], published in volume 406, *The Lancet*, which discusses the salient findings emerging from the pivotal IoN trial, led by Mallick et al.[2], published in the same volume of *The Lancet*. The escalating global incidence of thyroid carcinoma, with the preponderance of early-stage presentation being observed within high-income nations, renders the critical debate concerning potential overtreatment a matter of profound clinical and economic moment. As such, Hartl's[1] analysis rightly emphasizes the dramatic paradigm shift now underway in low-risk disease, characterized by tumors of stage T1 or T2, with nodal status N0 or Nx, where the recurrence rate is estimated at under 5% and disease-specific mortality remains below 1%. The IoN trial, a substantial prospective, randomized investigation including 504 patients, provides irrefutable, high-level confirmation that the post-operative omission of radioiodine ablation is demonstrably non-inferior to systematic treatment for this cohort, which finding significantly buttresses the conclusions drawn from the preceding ESTIMABL2 trial. The measured Kaplan-Meier 5-year estimated disease-free survival rate for the cases spared ablation was recorded at 97.9% (96.1–99.7), which compares most favorably with the 96.3% (93.9–98.7) observed in the ablation arm. Notably, no fatalities attributable to thyroid cancer were recorded across either cohort. Furthermore, the economic ramifications are compelling, suggesting a potential saving of 100 per patient within the UK's National Health Service, solely related to the costs of hospitalization and radioiodine administration for these procedures. Of the 504 patients randomly assigned, 251 were allocated to the no ablation group. Moreover, it remains incumbent upon the academic community to reflect upon the precise nuances elaborated on by Hartl[1], particularly regarding the broadened inclusion criteria of the IoN study. The trial admitted a minority of patients exhibiting T3 or T3a status (46 patients, or 9%) and N1a tumors (47 patients, or 9%). While this expansion beyond the confines of ESTIMABL2 is scientifically notable, the authors of the original Comment wisely caution against applying these definitive results to this intermediate subgroup, owing to the trial's lack of statistical power for rigorous subgroup analysis. Of note, despite the potential cost savings identified, a comprehensive, accurate cost-effectiveness analysis encompassing the entirety of the patient pathway remains unperformed. Likewise, the IoN trial was not designed to permit definitive conclusions regarding the predictive value of postoperative thyroglobulin concentrations for recurrence, meaning no such assertions may be drawn. The collective evidence, now profoundly reinforced by the IoN outcomes, serves as a crucial determinant in advancing the necessary individualization of therapy and mitigating the endemic issue of overtreatment in low-risk differentiated thyroid carcinoma. The logical progression, as articulated in the comment, suggests that proposing thyroid lobectomy for patients presenting with unilateral tumors may now be considered prudent. However, this shift mandates validating new surveillance algorithms, given the known unreliability of serum thyroglobulin as a predictor of tumor recurrence following lobectomy. Ultimately, a superior delineation of the underlying molecular architecture of these cancers is indispensable for reliably forecasting progression and recurrence, thereby ensuring that therapeutic intervention is reserved solely for those patients genuinely requiring it.

[1]Giresun University, Faculty of Medicine, Thyroidology Unit – Giresun, Turkey.
[2]Giresun University, Faculty of Medicine, Division of Endocrine Surgery – Giresun, Turkey.
[3]Giresun University, Faculty of Medicine, Department of General Surgery – Giresun, Turkey.
[4]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 December 16, 2025. Accepted on January 06, 2026.
Scientific Editor: José Maria Soares Júnior

This body of research provides a significant and necessary advancement in the clinical delineation of treatment pathways for differentiated thyroid cancer. This issue merits further investigation. We extend our sincere felicitation to Hartl[1] for their scholarly work in *The Lancet* on low-risk differentiated thyroid carcinoma for thyroidologists. *Felix qui potuit rerum cognoscere causas*.

ARTIFICIAL INTELLIGENCE

No generative artificial intelligence (AI) or AI-assisted technologies have been used in the creation and production of this article.

AUTHORS' CONTRIBUTIONS

IS: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing. **DS:** Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, 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. Hartl DM. Practice-changing evidence for low-risk differentiated thyroid cancer. Lancet. 2025;406:5-6.
2. Mallick U, Newbold K, Beasley M, Garcez K, Wadsley J, Johnson SJ, et al. Thyroidectomy with or without postoperative radioiodine for patients with low-risk differentiated thyroid cancer in the UK (IoN): a randomised, multicentre, non-inferiority trial. Lancet. 2025;406:52-62.
3. Sengul I, Sengul D. Construe with metabolic (dysfunction)-associated fatty liver disease, metabolic fatty liver disease, metrics, and contributions. Rev Assoc Med Bras. 2026;72(4):e202520709. https://orcid.org/10.1590/1806-9282.202520709
4. Sengul I, Sengul D. A consideration of the trends in thyroid fine-needle aspiration cytology – results from the Italian Cytopathology Committee National Practice Survey in thyroidology. Pathologica. 2026;118(1):49-50. https://orcid.org/10.32074/1591-951X-1852
5. Sengul D, Sengul I. Remarks on thyroid malignity in up-to-date category III of Novel Bethesda System for Reporting Thyroid Cytopathology in thyroidology. Rev Assoc Med Bras. 2025;71(10):e20250575. https://orcid.org/10.1590/1806-9282.20250575
6. Sengul I, Sengul D. Annotating risk stratification of thyroid nodules: assessing the suitability of ChatGPT for text-based analysis in thyroidology. Am J Otolaryngol. 2025;46(5):104696. https://orcid.org/10.1016/j.amjoto.2023.104144

https://doi.org/10.1590/1806-9282.20260994

Dermatological conditions in endurance athletes: prevalence and associated factors in a cross-sectional online survey

Gabriele Justino Paniago[1], Tatiana Cristina Figueira[1], Hélio Amante Miot[2]*

SUMMARY

OBJECTIVE: The aim of this study was to evaluate the prevalence of dermatological conditions among endurance athletes and investigate associations with demographic and sports-related variables.

METHODS: A cross-sectional online survey was conducted among 400 Brazilian adult endurance athletes, defined as individuals regularly training ≥10 h/week and/or participating in long-distance endurance events. Participants completed a structured questionnaire including demographic data, sports-related variables, sun exposure, photoprotection habits, and self-reported dermatological conditions. Associations between sports modalities and dermatological outcomes were assessed using logistic regression, with effect sizes expressed as odds ratios and 95%CIs.

RESULTS: Dermatological conditions were highly prevalent, with mechanically induced lesions representing the most frequent group. The most common conditions were foot blisters (73.8%), excoriations (59.0%), calluses (56.3%), abrasions (54.8%), and traumatic onycholysis (53.8%). First-degree sunburn (40.8%) and insect stings (36.3%) were the most frequent inflammatory/environmental conditions. Running was independently associated with foot blisters (OR 5.0; 95%CI 2.2–11.7) and traumatic onycholysis (OR 3.2; 95%CI 1.4–7.6), whereas cycling was associated with bacterial folliculitis (OR 2.7; 95%CI 1.0–7.5). Most athletes reported sun exposure exceeding 1 h per training session, while regular sunscreen use remained incomplete (56.4%). Self-reported skin neoplasms were uncommon but included seven melanoma cases, all among runners.

CONCLUSION: Endurance athletes exhibit a substantial burden of dermatological conditions related to repetitive mechanical stress, ultraviolet exposure, and sport-specific practices. Educational and preventive strategies focused on friction reduction, photoprotection, and early recognition of suspicious skin lesions may reduce short- and long-term morbidity.

KEYWORDS: Prevalence. Skin diseases. Athletes. Sports medicine. Cross-sectional studies.

INTRODUCTION

Outdoor exercise has increased worldwide in recent years, particularly since the COVID-19 pandemic, with running and cycling remaining among the most popular global fitness trends, thereby exposing athletes to ultraviolet radiation, water, environmental irritants, and repetitive mechanical trauma[1,2].

Endurance sports comprise prolonged aerobic activities such as long-distance running, cycling, triathlon, and open-water swimming, requiring sustained physical effort over extended periods[3]. In addition to the high physical demands, athletes are also exposed to environmental factors and mechanical friction[2]. Consequently, repetitive loading and environmental exposure may result in cumulative tissue stress affecting joints, muscles, skin, and cutaneous appendages[2,4].

Although skin disorders account for a substantial proportion of medical complaints in athletes, estimated at 10–20% in some sports medicine settings, epidemiological data specifically addressing endurance athletes remain limited[4]. Existing studies suggest that certain strategies, such as proper equipment care, adequate hygiene, sun protection, and dermatological follow-up, may reduce the risk of skin disorders. However, comprehensive epidemiological studies simultaneously evaluating multiple dermatological conditions across different endurance sport modalities remain scarce.

In this regard, it is important to highlight the main dermatological conditions affecting athletes, including mechanically induced skin disorders, skin infections, inflammatory conditions, and skin cancer. Mechanically induced skin diseases are among the most common and may negatively affect both training and performance. Repetitive skin trauma can alter skin structure and lead to various clinical manifestations. Skin infections are also common in sports due to athlete-to-athlete contact, inadequate equipment hygiene, environmental exposure, and trauma. Mechanical lesions such as friction blisters and traumatic nail

[1]Universidade Estadual Paulista, Faculdade de Medicina de Botucatu – Botucatu (SP), Brazil.
[2]Universidade Estadual Paulista, Faculdade de Medicina de Botucatu, Department of Dermatology and Radiotherapy – Botucatu (SP), Brazil.
*Corresponding author: heliomiot@gmail.com

Conflicts of interest: the authors declare there is no conflicts of interest. Funding: this study was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil (PIBIC grant no. 128096/2024-1; Research Productivity grant no. 306358/2022-0).
Received on June 16, 2026. Accepted on July 20, 2026.
Scientific Editor: Roseli Nomura

changes are among the most common dermatological disorders in runners, whereas moisture-related bacterial and fungal dermatoses are particularly relevant in athletes exposed to prolonged sweating and occlusion. Moreover, inflammatory conditions such as urticaria and contact dermatitis are particularly relevant in sports settings, as their management requires knowledge of both equipment and environmental conditions[2,5]. Skin cancer is another important concern, as athletes are often exposed to high doses of ultraviolet radiation due to prolonged outdoor training schedules[5-7]. Furthermore, sports activities may aggravate pre-existing dermatological conditions, which may or may not be previously known to the athlete[5].

International studies have shown that outdoor athletes frequently sustain sunburns, receive ultraviolet doses well above occupational exposure thresholds, and often report incomplete adherence to protective measures. Studies from the SOLUDABLE Research Group in Spain have documented high rates of sunburn and inadequate sun protection in water-sport athletes and trail runners, while an Australian study has reported substantial burdens of melanoma and keratinocyte cancer among people engaging in outdoor sports and recreation. Endurance athletes are repeatedly exposed to ultraviolet radiation during training and competitions, often under conditions that reduce the effectiveness of photoprotection, such as sweating, friction, water exposure, and prolonged exercise duration[6-10].

Dermatologic literature on sports in Brazil (and Latin America) is scarce, and to date, no nationwide study has evaluated the prevalence of dermatological conditions across multiple endurance sports in Brazilian athletes.

This study aimed to assess the prevalence of reported dermatological conditions among endurance athletes and investigate associations with sport modality and demographic characteristics.

METHODS

This cross-sectional observational study was conducted as an Internet-based survey among Brazilian adult (≥18 years) endurance athletes. Endurance athletes were defined as individuals regularly practicing long-duration aerobic sports (running, cycling, swimming, or triathlon), with weekly training ≥10 h and/or participation in long-distance competitions. Participants were recruited through social media platforms (Facebook, Instagram, and WhatsApp). Incomplete questionnaires and individuals whose sports practice did not meet the definition of endurance sports were excluded.

Data were collected using a structured online questionnaire developed in Google Forms, including demographic characteristics, sports-related variables (sport modality, training intensity, and weekly training duration), and self-reported dermatological conditions associated with sports practice. To improve the accuracy of self-reported data and reduce misclassification bias, illustrative images of common dermatological conditions were included to assist lesion identification. The questionnaire was developed specifically for this study based on the sports dermatology literature and expert consensus. Before dissemination, it underwent face review by two dermatologists experienced in sports dermatology to improve clarity and comprehensibility. Before dissemination, the questionnaire was pilot-tested in 10 endurance athletes to assess clarity and comprehensibility, without the need for adjustments. Formal psychometric validation was not performed. Google Forms was selected because it allows anonymous participation, secure institutional data storage, restricted investigator access, and encrypted transmission.

The primary outcome was the frequency of self-reported dermatological conditions. Secondary outcomes included demographic characteristics, sports-related variables, and patterns of sun exposure and photoprotection. Descriptive statistics, including percentages and measures of central tendency and dispersion, were used to summarize the data. Effect sizes were expressed as multivariate odds ratios (OR), adjusted for age, sex, and weekly training volume. Statistical significance was set at $p<0.05$[11].

Although the study used convenience sampling and was not designed for probabilistic inference, the target of approximately 400 completed questionnaires was pragmatically defined to obtain reasonably stable prevalence estimates in a nationwide exploratory survey[12]. Analyses were performed using IBM SPSS 31v.

This study was conducted in accordance with the ethical standards of the institutional research ethics committee. The protocol was approved by the Research Ethics Committee of the Botucatu School of Medicine (CEP-FMB/UNESP) under approval number 6,964,765 (CAAE: 79996324.8.0000.5411). Participation was voluntary, and informed consent was obtained electronically. No personally identifiable information was collected, ensuring anonymity and confidentiality.

RESULTS

A total of 400 valid questionnaires from endurance athletes were analyzed (Figure 1), with demographic, exposure, and sports-related characteristics presented in Table 1. Most participants were male, aged 30–50 years, predominantly fair-skinned (self-reported White), and highly educated. Running was the most common modality, and most participants trained more than 10 h per week.

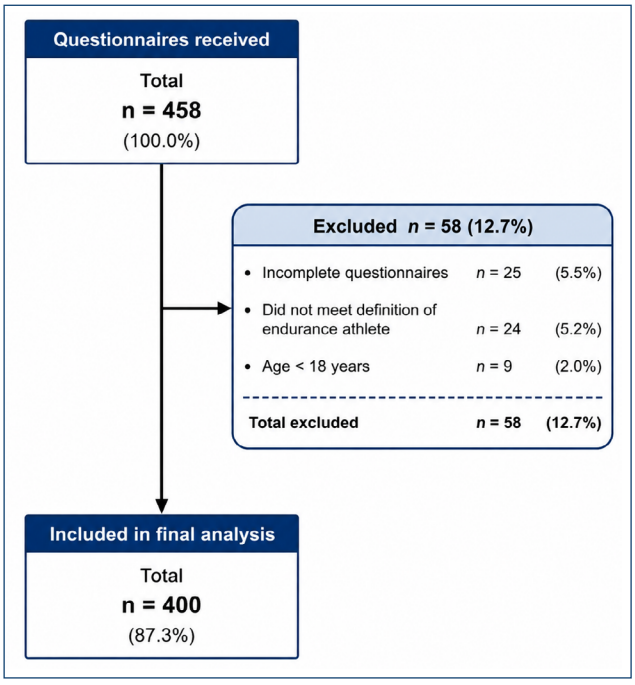


Figure 1. Flow diagram of participant selection.

Sun exposure was highly prevalent, with most athletes reporting exposure exceeding 1 h per training session. Although sun protection measures were commonly reported, adherence to regular sunscreen use was incomplete.

Dermatological conditions were highly prevalent in this population (Table 2), with mechanical injuries representing the most frequently reported group. Mechanical lesions predominated, affecting more than half of the participants. Foot blisters were reported by 73.8%, excoriations by 59.0%, calluses by 56.3%, abrasions by 54.8%, and traumatic onycholysis by 53.8%. The most frequent inflammatory conditions were sunburn and insect bites. Notably, seven participants reported a previous diagnosis of melanoma, three reported basal cell carcinoma, and one reported squamous cell carcinoma. Notably, jellyfish stings were reported in 21.6% of open-water swimmers, highlighting sport-specific environmental exposure.

Table 3 presents the association between sports modality and the main reported injuries. Cycling was associated with bacterial folliculitis, insect bites, abrasions, and lip fissures. Running was linked to foot blisters and traumatic onycholysis. Regarding demographic factors, younger athletes were more likely to report foot blisters (mean age 40 vs. 46 years, p<0.01). Nipple irritation was strongly associated with male athletes (OR 20.8; 95%CI 8.2–52.5; p<0.01), whereas female athletes were more likely to adopt protective behaviors, including the use of sunscreen (OR 15.5; 95%CI 8.0–30.0; p<0.01) and moisturizers (OR 6.6; 95%CI 4.3–10.3; p<0.01).

Table 1. Demographic, exposure, and endurance-related characteristics (n=400).

Variables	Values	
Sex, n (%)		
Female	163	(40.8%)
Male	237	(59.3%)
Age (in years)		
Mean (SD)	41.5	(10.2)
Maximum	73	
Minimum	18	
Self-declared skin color, n (%)		
White	305	(76.4%)
Brown	69	(17.3%)
Black	12	(3.0%)
Asian	11	(2.8%)
Indigenous	2	(0.5%)
Education level, n (%)		
Elementary	3	(0.8%)
High school	28	(7.0%)
Higher education	369	(92.3%)
Sports modalities, n (%)		
Running	360	(90.0%)
Cycling	255	(63.8%)
Swimming	190	(47.5%)
Duration of sports practice (in years)		
Mean (SD)	9.11	(7.8)
Maximum	40	
Minimum	0.3	
Weekly training time, n (%)		
Up to 10 h	37	(9.3%)
11–15 h	206	(51.5%)
More than 15 h	157	(39.3%)
Frequency of sun exposure during sports practice per week, n (%)		
1 to 3 times	211	(52.8%)
4 to 6 times	155	(38.8%)
Every day	33	(8.3%)
Sun exposure duration during sports practice, n (%)		
1 h	252	(63.0%)
2–3 h	42	(10.5%)
More than 4 h	106	(26.5%)
Sun protection measures during outdoor sports practice	262	(65.8%)
Regular use of sunscreen for sports	225	(56.4%)
Use of sun-protective clothing	284	(71.4%)
Use of a cap/hat for sports	316	(79.2%)
Use of sunglasses for sports	175	(44.0%)

SD: standard deviation.

DISCUSSION

This study provides evidence that dermatological conditions were highly prevalent among Brazilian endurance athletes, and different endurance sports were associated with specific dermatological risks, reflecting the combined effects of environmental exposure and mechanical stress.

Table 2. Main dermatological conditions reported by endurance athletes (n=400).

Variables	Results	
	n	%
Mechanically induced lesions		
Foot blisters	295	73.8
Cuts/excoriations	236	59.0
Calluses on hands or feet	225	56.3
Abrasion on the thigh/groin	219	54.8
Onycholysis after sports activity	215	53.8
Nipple irritation	99	24.8
Mechanical acne	59	14.8
Piezogenic papules	6	1.5
Inflammatory and infectious conditions		
First-degree sunburn	163	40.8
Insect/wasp sting	145	36.3
Lip fissures	131	32.8
Interdigital intertrigo	104	26.0
Herpes simplex	77	19.3
Heat/sweat-induced urticaria	64	16.0
Abscesses	61	15.3
Bacterial folliculitis	54	13.5
Tinea pedis	39	9.8
Heat rash (“prickly heat”)	32	8.0
Tinea cruris	30	7.5
Sun-induced urticaria	28	7.0
Sunscreen allergy	26	6.5
Jellyfish sting (marine exposure)	24	6.0
Cold-induced urticaria	16	4.0
Clothing/sportswear allergy	10	2.5
Plantar keratolysis	8	2.0
Bacterial skin infection (impetigo)	5	1.3
Second-degree sunburn	2	0.5
Molluscum contagiosum	1	0.3
Skin neoplasias		
Melanoma	7	1.8
Basal cell carcinoma	3	0.8
Squamous cell carcinoma	1	0.3

Mechanical skin injuries were the most prevalent conditions, particularly among runners and cyclists. Excoriations and folliculitis, commonly reported by cyclists, may be related to friction and prolonged contact with equipment, and could potentially be prevented with friction-reduction strategies and barrier creams. While female athletes often use compression garments or sports bras, nipple protection appears particularly relevant for male runners. Nipple chafing has been reported in up to one-third of long-distance runners, especially men, due to repetitive shirt friction[2,13]. These findings suggest that both biological and behavioral factors contribute to the distribution of dermatological conditions in endurance athletes.

Mechanical lesions were the dominant group in our study, especially among runners, who showed higher odds of foot blisters and post-sport onycholysis. This pattern is biologically plausible and agrees with the sports dermatology literature, in which repetitive shear forces, moisture, footwear characteristics, and prolonged distance are major determinants of frictional injury[8].

Prolonged aerobic activity induces sweating, which alters stratum corneum hydration and reduces the minimal erythema dose[14]. In addition, endorphin release during exercise may decrease the perception of pain and sunburn, potentially increasing the risk of unnoticed skin injury. Together, these mechanisms may contribute to the higher susceptibility of endurance athletes to dermatological conditions.

Brazil presents one of the world’s highest ambient UV indices, with year-round outdoor sports participation. Consequently, Brazilian endurance athletes may experience different patterns of cumulative ultraviolet exposure compared with athletes from temperate climates, reinforcing the relevance of locally generated epidemiological data. Thus, sun exposure was a relevant factor in this study.

Photoprotection remains a challenge for outdoor athletes due to both reduced sunscreen effectiveness under conditions of sweating, friction, and repeated washing and low adherence, particularly among male athletes. Previous studies suggest that only 39–66% of outdoor athletes adequately use or reapply sunscreen, with consistently poorer adherence among men[15,16]. Proper education of athletes, coaches, and healthcare professionals regarding the use of broad-spectrum sunscreens, sun-protective clothing, and training during periods of lower ultraviolet radiation is essential to reduce both acute and long-term skin damage.

Although skin neoplasms were reported in a small subset of participants, their presence is clinically relevant. All seven self-reported melanoma cases occurred in runners aged 40–56 years. Among these individuals, 71.4% reported sunscreen use, whereas only 28.6% used sun-protective clothing. Despite this

Table 3. Multivariate association between the main types of dermatological lesions and the sports modality. Odds ratios adjusted for age, sex, and weekly training volume.

Inflammatory and infectious conditions	Cycling	Swimming	Running
Abscesses	1.6 (0.6–4.1)	1.8 (0.7–4.3)	0.9 (0.3–2.6)
Bacterial folliculitis	**2.7 (1.0–7.5)***	1.1 (0.5–2.8)	0.5 (0.2–1.4)
Herpes simplex	1.8 (0.8–4.2)	0.8 (0.4–1.8)	3.0 (0.9–10.4)
Fungal infection between the toes	1.2 (0.6–2.6)	1.4 (0.7–2.9)	1.1 (0.4–2.9)
Insect sting/wasp	**2.1 (1.1–4.2)***	1.0 (0.5–1.9)	0.7 (0.3–1.5)
First-degree sunburn	1.5 (0.8–3.0)	1.5 (0.8–2.8)	1.0 (0.4–2.2)
Lip fissures	**2.1 (1.0–4.2)***	0.9 (0.4–1.7)	1.4 (0.6–3.3)
Heat/sweat urticaria	0.9 (0.4–2.1)	1.1 (0.4–2.6)	1.7 (0.5–5.8)
Mechanically induced lesions			
Abrasion on the thigh/groin	1.8 (0.9–3.5)	1.0 (0.5–1.9)	1.5 (0.7–3.4)
Blisters on the feet	0.8 (0.4–1.6)	1.6 (0.8–3.2)	**5.0 (2.2–11.7)***
Calluses on the hands or feet	0.7 (0.3–1.3)	1.4 (0.8–2.7)	2.0 (0.9–4.5)
Nail detachment after sports	0.9 (0.5–1.7)	1.1 (0.6–2.1)	**3.2 (1.4–7.6)***
Nipple irritation	1.5 (0.7–3.1)	0.6 (0.3–1.3)	1.9 (0.7–5.1)
Cuts/excoriations	**4.6 (2.3–9.3)***	0.8 (0.4–1.6)	0.7 (0.3–1.7)

*Bold values: p<0.05.

clustering among runners, no inference regarding incidence or causality can be drawn from these findings. Nevertheless, previous studies have identified marathon runners as a population at increased risk for melanoma[17]. In accordance with international literature, our findings suggest that inadequate photoprotection is consistently observed across different endurance sports[6,7,16].

Our findings are consistent with the growing international literature showing that outdoor athletes are exposed to substantial dermatological morbidity related to both friction and solar ultraviolet radiation. Studies from different countries reported very high rates of sunburn and inadequate sun protection in elite water-sport athletes and trail runners. In water sports, 76.7% of athletes reported sunburn during the previous season, whereas in high-altitude trail running, 66.5% reported sunburn and nearly half did not reapply sunscreen. These figures are consistent with the incomplete sunscreen adherence observed in our sample, despite frequent outdoor exposure[6,7,9,15,16].

Australian research provides additional support for the long-term relevance of these behaviors. In a recent Queensland study combining questionnaires with full-body skin examination and histopathologic confirmation, melanoma prevalence was significantly higher in aquatic than in land-based outdoor participants. Similarly, an Australian study of surfers and swimmers has reported substantial lifetime burdens of skin cancer. Together, these findings place our Brazilian data within an international pattern in which repeated outdoor training is associated with considerable acute photodamage and may contribute to long-term skin cancer burden[8].

The higher occurrence of skin cancer among runners highlights the need for targeted prevention and early detection strategies. Outdoor endurance athletes may accumulate ultraviolet doses two to eight times higher than indoor populations, reinforcing the biological plausibility of increased photocarcinogenic risk[9,10,18]. The observed frequency should not be interpreted as prevalence in the athletic population because participants constituted a convenience sample with possible self-selection of individuals interested in dermatologic diseases.

Because participants were recruited through social media using convenience sampling, the findings should not be interpreted as nationally representative of all Brazilian endurance athletes. However, the study provides one of the largest multicenter descriptions currently available for this population. Moreover, the questionnaire was developed for this study and did not undergo formal psychometric validation, which may have affected the reliability and construct validity of some measures. All dermatological outcomes were self-reported, without clinical confirmation, which may have introduced non-differential or differential misclassification, particularly for infectious and neoplastic conditions. Third, social media recruitment yielded a convenience sample, limiting generalizability and potentially favoring highly engaged, highly educated athletes.

Because several dermatological outcomes were highly prevalent, prevalence ratios might have provided more intuitive measures of association than OR. Nevertheless, OR remain a valid and widely accepted approach for exploratory cross-sectional

analyses, particularly when the objective is to identify independent associations. Moreover, given the cross-sectional design, all regression estimates should be interpreted as measures of association rather than evidence of causal relationships.

From a practical perspective, these findings support targeted educational interventions for athletes, coaches, and sports health professionals. Preventive strategies should emphasize friction control (appropriate footwear, moisture management, moisture-wicking socks, and clothing selection), regular inspection of feet and nails, and structured photoprotection, including broad-spectrum sunscreen application 15–30 min before exercise, reapplication during prolonged training, use of UV-protective clothing, and preferential scheduling of outdoor sessions outside peak ultraviolet hours. These simple measures may substantially reduce both acute skin injuries and long-term photocarcinogenic risk. Future prospective studies using clinically confirmed diagnoses are warranted to validate these findings.

CONCLUSION

Endurance athletes exhibit a substantial burden of dermatological conditions associated with mechanical stress, ultraviolet exposure, and sport-specific practices. These findings reinforce the importance of educational strategies for athletes and coaches, focusing on photoprotection, friction reduction, equipment optimization, and early recognition of suspicious skin lesions to reduce short- and long-term morbidity.

ETHICS APPROVAL

This study was approved by the Research Ethics Committee of Faculdade de Medicina de Botucatu, under approval number 6.964.765 (CAAE: 79996324.8.0).

AUTHORS' CONTRIBUTIONS

GJP: Conceptualization, Data curation, Formal Analysis, Writing – original draft, Writing – review & editing. **TCF:** Conceptualization, Data curation, Formal Analysis, Writing – original draft, Writing – review & editing. **HAM:** 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. Thompson WR. Worldwide survey of fitness trends for 2025. ACSMs Health Fit J. 2024;28:14-23.
2. Carr PC, Cropley TG. Sports dermatology: skin disease in athletes. Clin Sports Med. 2019;38:597-618.
3. Matomäki P. Why low-intensity endurance training for athletes? Eur J Appl Physiol. 2025;125:2401-7.
4. Adams BB. Sports dermatology. Adolesc Med. 2004;15:403-13.
5. Freiman A, Barankin B, Elpern DJ. Sports dermatology part 1: common dermatoses. CMAJ. 2004;171:851-3.
6. Castro-Maqueda G, Gutierrez-Manzanedo JV, Lagares-Franco C, Troya-Martin M. Sun exposure during water sports: do elite athletes adequately protect their skin against skin cancer? Int J Environ Res Public Health. 2021;18:800.
7. García-Malinis AJ, Gracia-Cazaña T, Zazo M, Aguilera J, Rivas-Ruiz F, Troya Martín M, et al. Sun protection behaviors and knowledge in mountain marathon runners and risk factors for sunburn. Actas Dermosifiliogr (Engl Ed). 2021;112:159-66.
8. Climstein M, Doyle B, Stapelberg M, Rosic N, Hertess I, Furness J, et al. Point prevalence of non-melanoma and melanoma skin cancers in Australian surfers and swimmers in Southeast Queensland and Northern New South Wales. PeerJ. 2022;10:e13243.
9. Kliniec K, Tota M, Zalesińska A, Łyko M, Jankowska-Konsur A. Skin cancer risk, sun-protection knowledge and behavior in athletes: a narrative review. Cancers (Basel). 2023;15:3281.
10. Moehrle M. Outdoor sports and skin cancer. Clin Dermatol. 2008;26:12-5.
11. Miola AC, Miot HA. P-value and effect-size in clinical and experimental studies. J Vasc Bras. 2021;20:e20210038.
12. Miot HA. Sample size in clinical and experimental trials. J Vasc Bras. 2011;10(4):275-8.
13. Luca JF, Adams BB, Yosipovitch G. Skin manifestations of athletes competing in the summer olympics: what a sports medicine physician should know. Sports Med. 2012;42:399-413.
14. Moehrle M, Koehle W, Dietz K, Lischka G. Reduction of minimal erythema dose by sweating. Photodermatol Photoimmunol Photomed. 2000;16:260-2.
15. Sesi J, Yousif J, Musa A, Warbasse E, Cohen G. Skin in the game: inadequate photoprotection among Olympic athletes. Cutis. 2023;112:E38-39.
16. Fernandez-Ruiz J, Montero-Vilchez T, Buendia-Eisman A, Arias-Santiago S. Knowledge, behaviour and attitudes related to sun exposure in sportspeople: a systematic review. Int J Environ Res Public Health. 2022;19:10175.
17. Ambros-Rudolph CM, Hofmann-Wellenhof R, Richtig E, Müller-Fürstner M, Soyer HP, Kerl H. Malignant melanoma in marathon runners. Arch Dermatol. 2006;142(11):1471-4.
18. Moehrle M, Heinrich L, Schmid A, Garbe C. Extreme UV exposure of professional cyclists. Dermatology. 2000;201(1):44-5.

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20251923

Do hospitalized children with obesity lose more weight than eutrophic ones?

Dalila Nunes Alcantara[1], Karina Guimarães de Angelis[1], Adriana Sañudo[2], Maria Teresa Bechere Fernandes[1]*

SUMMARY

BACKGROUND: The clinical course of hospitalized children with overweight has been scarcely investigated. Previous studies suggest that these children experience longer hospital stays and greater weight loss than their normal-weight peers, potentially compromising lean body mass and immune function.

OBJECTIVE: The aim of this study was to investigate the association between nutritional status at hospital admission and hospital length of stay, and to compare weight loss trajectories between children with overweight and those with normal weight.

METHODS: This prospective study included 125 hospitalized children and adolescents aged 2–14 years. Anthropometric measurements and sociodemographic and clinical data were collected. Associations with hospital length of stay were assessed using multiple Poisson regression, and weight loss trajectories were analyzed using analysis of covariance.

RESULTS: Of the 125 participants, 56.8% were male, and 29.6% had overweight. Asthma was the most common reason for hospitalization (25.6%). Children with overweight had a 25% longer hospital stay than those with normal weight. Among children with overweight, each additional day of hospitalization was associated with a mean weight loss of 0.20 kg (95%CI 0.14–0.26) and a 0.04-unit decrease in body mass index-for-age z-score (95%CI 0.01–0.06).

CONCLUSION: Overweight at hospital admission was associated with a longer hospital stay and progressive weight loss during hospitalization.

KEYWORDS: Teenagers. Children. Hospital. Hospitalization. Overweight. Weight loss.

INTRODUCTION

Pediatric overweight and obesity have reached alarming levels worldwide, including in Brazil. A nationwide study including more than five million Brazilian children aged 3–10 years reported an increase in the prevalence of excess weight of approximately 3% among those aged 5–10 years and 1% among those aged 3–4 years between 2001 and 2014[1].

Although hospital malnutrition has been extensively investigated, the impact of excess weight on hospitalization outcomes remains poorly understood. In Brazil, the prevalence of overweight among hospitalized children ranges from 8.9 to 25.2%[2,3], while international studies have reported prevalences ranging from 17.3 to 30%[4-6].

A systematic review published in 2013 identified childhood obesity as a risk factor for increased mortality in severe illnesses[7]. In addition, children with obesity are more likely to require hospitalization for orthopedic, respiratory, surgical, gastrointestinal, and genitourinary conditions, resulting in increased healthcare utilization and hospital costs[4,8].

Excess weight has also been consistently associated with asthma. A 2019 meta-analysis including 73,252 children reported a higher risk of asthma among children with overweight (OR 1.30; 95%CI 1.23–1.39)[9], and a subsequent meta-analysis confirmed this association (OR 1.50; 95%CI 1.30–1.70)[10]. Likewise, a retrospective cohort study found that children with overweight were more likely to be hospitalized because of asthma (OR 1.60; 95%CI 1.00–2.60)[11].

Potential mechanisms include obesity-related immune dysregulation and chronic low-grade inflammation, which may contribute to poorer clinical outcomes and prolonged hospital stays[2,12,13].

Given the limited evidence on the clinical course of hospitalized children with overweight, this study aimed to investigate the association between nutritional status at hospital admission and hospital length of stay and to compare weight

[1]Universidade de São Paulo, Faculdade de Medicina, Department of Pediatrics – São Paulo (SP), Brazil.
[2]Universidade Federal de São Paulo, Department of Preventive Medicine – São Paulo (SP), Brazil.
*Corresponding author: maria.teresa@fm.usp.br
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: CAPES Scholarship and USP Unified Scholarship Program Scholarship.
Received on May 21, 2026. Accepted on August 14, 2026.
Scientific Editor: Roseli Nomura

loss trajectories during hospitalization between children with overweight and those with normal weight.

METHODS

Study design and participants

This prospective cohort study included 125 children and adolescents aged 2–14 years who were admitted to the University Hospital of the Universidade de São Paulo, Brazil, between 2014 and 2017. Only the first hospitalization during the study period was considered for inclusion. Children younger than 2 years of age or with chronic diseases were excluded.

The sample size was estimated using G*Power version 3.1, assuming mean hospital stays of 5.5 and 3.7 days for children with overweight and normal weight, respectively, a common standard deviation (SD) of 2.6 days, a two-sided α of 0.05, and 80% statistical power, resulting in a minimum sample size of 33 participants per group[2]. To ensure adequate statistical power, 88 children with normal weight and 37 with overweight were included.

Data collection and variables

Sociodemographic information, including child and maternal characteristics and family income, was obtained through structured interviews. Family income was expressed as multiples of the Brazilian minimum wage. Clinical data, including the primary reason for hospitalization, antibiotic use, oxygen therapy, and admission to the intensive care unit (ICU), were obtained from medical records.

Anthropometric measurements were performed by trained researchers within 24 h of hospital admission. Body weight was measured daily using a calibrated FILIZOLA® digital scale, and height was measured using an ALTURAEXATA® stadiometer according to standardized procedures[14,15]. Waist circumference was measured at the midpoint between the lowest rib and the iliac crest.

Nutritional status was classified according to the World Health Organization (WHO) growth standards[16] using WHO Anthro version 3.2.2 and WHO AnthroPlus version 1.0.4 to calculate body mass index (BMI)-for-age (BAZ) and height-for-age (HAZ) z-scores. The overweight group comprised children classified as at risk of overweight, overweight, or obese. Central adiposity was assessed using the waist-to-height ratio, with values ≥0.50 indicating abdominal adiposity[17].

Statistical analysis

Categorical variables are presented as frequencies and continuous variables as mean (SD) or median (interquartile range), as appropriate. Because hospital length of stay was not normally distributed, the variable was log-transformed and is presented as the geometric mean with a 95%CI.

Factors associated with hospital length of stay were evaluated using Poisson regression models. Variables with $p<0.20$ in the univariate analyses were included in the multivariable model. Results are presented as relative risks with corresponding 95%CI.

Pearson correlation coefficients were used to examine the associations between hospital length of stay and changes in body weight (Δweight) and BMI-for-age z-score (ΔBAZ). Delta values were calculated as the difference between discharge and admission measurements, with negative values indicating weight loss.

Changes in body weight and ΔBAZ were further analyzed using analysis of covariance (ANCOVA), with nutritional status as the fixed factor and hospital length of stay as a covariate. Interaction terms were included to evaluate whether the association between hospital length of stay and anthropometric changes differed according to nutritional status. Statistical significance was defined as a two-sided $p<0.05$. All analyses were performed using STATA/MP version 18.0 (StataCorp LLC, College Station, TX, USA).

Ethical considerations

The study was approved by the Research Ethics Committee of the Faculty of Medicine and the University Hospital of the Universidade de São Paulo (CAAE: 27944414.6.0000.0065). Written informed consent was obtained from the parents or legal guardians of all participants before enrollment.

RESULTS

The study included 125 hospitalized children and adolescents with a mean age of 7.4 years, of whom 56.8% were male. Mothers had a mean of 9.6 years of schooling, and the median family income was 2.1 times the Brazilian minimum wage. At hospital admission, 70.4% of participants had normal weight and 29.6% had overweight. The mean WHtR was 0.50, indicating central adiposity in a substantial proportion of the sample. The most common reasons for hospitalization were asthma (25.6%), cellulitis (14.4%), and appendicitis (12.8%). The median hospital length of stay was 4 days. Overall, 68.0% of patients received antibiotics, 22.6% received oxygen therapy, and 6.5% required admission to the ICU (Table 1).

In the multivariable Poisson regression analysis, overweight status, antibiotic use, and ICU admission were independently associated with longer hospital stays ($p<0.05$), whereas oxygen

Table 1. Description of the sample according to sociodemographic and clinical data (n=125).

	Mean (SD)/median (P25–P75) Minimum–maximum	n (%)
Age (years)	7.4 (3.4)	
	2.4–14.9	
Sex		
Female		54 (43.2)
Male		71 (56.8)
Maternal age (years)	32.6 (7.5)	
	18.0–52.0	
Education (years)	9.6 (3.0)	
	1–16	
Income (minimum wage)	2.1 (1.4–3.8)	
	0.4–12.7	
Nutritional status		
Eutrophy		88 (70.4)
Risk of overweight		4 (3.2)
Overweight		13 (10.4)
Obesity		20 (16.0)
HAZ (z-escore)	0.2 (1.3)	
	-2.6–3.2	
BAZ (z-escore)	0.5 (1.3)	
	-2.0–3.6	
Waist (cm)	58.5 (52.0–66.0)	
	43.0–100.0	
Waist-to-height ratio	0.5 (0.1)	
	0.4–0.7	
Diagnosis at the time of hospitalization		
Abscess		3 (2.4)
Appendicitis		16 (12.8)
Acute appendicitis with abscess		7 (5.6)
Septic arthritis		2 (1.6)
Asthma		32 (25.6)
Asthma and pneumonia		11 (8.8)
Acute bronchiolitis		2 (1.6)
Cellulitis		18 (14.4)
Other infectious diseases		13 (10.4)
Diabetes mellitus–insulin-dependent		5 (4.0)
Fracture		3 (2.4)
Diseases of the nervous system		3 (2.4)
Others		10 (8.0)
Length of hospitalization (days)	4 (2–7)	
	1–36	
Use of antibiotics		
No		40 (32.3)
Yes		84 (67.7)
Use of oxygen		
No		96 (77.4)
Yes		28 (22.6)
ICU		
No		115 (93.5)
Yes		8 (6.5)

SD: standard deviation; HAZ: height-for-age z-score; BAZ: body mass index-for-age z-score; ICU: intensive care unit.

therapy was associated with a shorter length of stay (p<0.001) (Table 2). Children with overweight experienced greater reductions in body weight (Δweight=-0.09±0.36 kg) and BMI-for-age z-score (ΔBAZ=-0.09±0.36) than children with normal weight (Δweight=-0.03±0.43 kg).

Among children with normal weight, hospital length of stay was not significantly correlated with changes in body weight or BMI-for-age z-score. In contrast, among children with overweight, longer hospital stays were significantly associated with greater reductions in both body weight (r=-0.656) and BMI-for-age z-score (r=-0.540). ANCOVA confirmed a significant interaction between nutritional status and hospital length of stay for both Δweight (interaction β=-0.23) and ΔBAZ (interaction β=-0.05; both p<0.01). Specifically, among children with overweight, each additional day of hospitalization was associated with an average weight loss of 0.20 kg and a 0.04-unit decrease in BMI-for-age z-score (Figure 1).

DISCUSSION

This study showed that children with overweight had a 25% longer hospital stay than children with normal weight and experienced progressive, time-dependent weight loss during hospitalization.

Between 2014 and 2017, 29.6% of hospitalized children in our cohort had overweight, corresponding to an average annual prevalence of 7.4%. This prevalence was comparable to that reported in a previous study conducted at the same institution 5 years earlier (8.9% annually)[2], but lower than the 25.2% observed at a tertiary referral hospital in Recife[3]. Differences in age eligibility, particularly the inclusion of infants younger than 1 year in the Recife study, may partly explain this discrepancy. International studies have reported prevalences ranging from 17.3 to 30% among hospitalized children and adolescents[4,6]. It is important to emphasize that these estimates represent the prevalence of overweight among hospitalized patients rather than in the general pediatric population. Consequently, they reflect a subgroup of children requiring hospital care and should not be directly compared with population-based prevalence estimates.

Asthma, cellulitis, and appendicitis were the leading causes of hospitalization in the present study. Asthma was also the most frequent diagnosis in a previous study conducted at our institution[2], and respiratory diseases remain one of the leading causes of pediatric hospitalization worldwide[6,18].

Overweight, antibiotic use, and ICU admission were independently associated with longer hospital stays. Antibiotic therapy and ICU admission are likely markers of

Table 2. Multiple Poisson regression model for length of hospitalization.

Multiple Poisson regression*	$RR_{adjusted}$	95%CI	p
Nutritional status			
Weight excess×eutrophy	1.25	1.02–1.55	0.036
Use of antibiotics (yes×no)	1.70	1.24–2.34	0.001
Use of oxygen (yes×no)	0.54	0.40–0.72	<0.001
ICU (yes×no)	2.21	1.52–3.22	<0.001

*Model adjusted for child's age, hospitalization diagnosis, maternal education, and income. CI: confidence interval; ICU: intensive care unit.

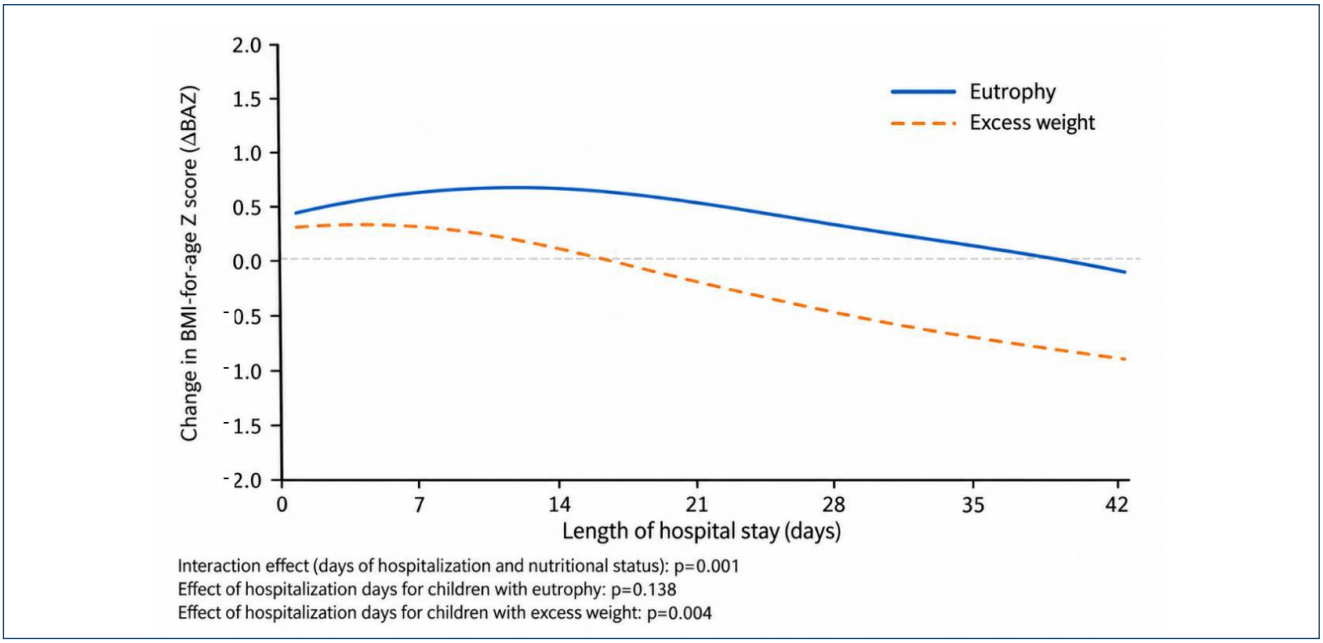


Figure 1. Interaction effect—days of hospitalization and nutritional status.

greater disease severity and clinical complexity. Conversely, oxygen therapy was associated with a shorter hospital stay, although this finding may reflect differences in clinical management or patient characteristics rather than a direct therapeutic effect.

Although overweight remained independently associated with prolonged hospitalization, hospital length of stay in pediatric patients should be viewed as a multidimensional outcome rather than a direct consequence of disease severity or nutritional status alone. Length of stay reflects the interaction of disease severity, recovery trajectory, care complexity, and healthcare delivery processes. Consistent with this perspective, a recent study demonstrated that nursing care demands independently contributed to prolonged hospitalization among children and adolescents, even after adjustment for diagnosis-related group severity and other clinical characteristics[19]. Therefore, the longer hospital stays observed among children with overweight may also reflect greater care complexity and evolving supportive care needs during hospitalization, rather than exclusively the biological effects of excess adiposity.

Our findings are consistent with previous pediatric studies showing that overweight and obesity are associated with prolonged hospitalization across a range of clinical settings, including respiratory diseases, surgical admissions, and COVID-19[4,6,20-22]. Together, these findings support the hypothesis that excess adiposity is associated with less favorable hospital outcomes.

The most important finding of the present study was the progressive, time-dependent weight loss observed among children with overweight during hospitalization. Each additional day of hospitalization was associated with an average reduction of 0.20 kg in body weight and a 0.04-unit decrease in BMI-for-age z-score. No similar pattern was observed among children with normal weight. To our knowledge, few pediatric studies have examined longitudinal changes in anthropometric indicators throughout hospitalization. A previous study conducted at our institution similarly reported greater weight z-score loss

among hospitalized children with overweight than among those with normal weight[2].

Several mechanisms may contribute to these findings. During acute illness, surgery, or trauma, children with overweight may develop a hypercatabolic response characterized by disproportionate loss of lean body mass relative to fat mass[23]. This metabolic response may impair immune function and contribute to poorer clinical outcomes in children with obesity, particularly those hospitalized with respiratory conditions such as asthma[10-12,24]. Nevertheless, these mechanisms remain speculative in the context of the present study and warrant further investigation.

The observed association between prolonged hospitalization and progressive weight loss should also be interpreted with caution. The overweight group included children ranging from at risk of overweight to obesity, representing a clinically heterogeneous population. Moreover, because body composition was not assessed, it was not possible to determine whether the observed weight loss reflected reductions in fat mass, lean mass, or body water. Future prospective studies incorporating body composition assessment and metabolic biomarkers are needed to better characterize the mechanisms underlying weight loss during hospitalization in children with overweight.

This study has some limitations. First, the sample size was relatively small (n=125) compared with the total number of pediatric hospitalizations during the study period (n=2,745). However, comparisons between included and non-included patients suggested that selection bias was minimal. Second, this was a single-center study, which may limit the generalizability of the findings to other healthcare settings.

CONCLUSION

Overweight at hospital admission was independently associated with a longer hospital stay and progressive, time-dependent weight loss during hospitalization. These findings suggest that children with overweight may represent a subgroup at increased nutritional vulnerability during hospitalization and highlight the importance of careful nutritional monitoring throughout the hospital stay. Further prospective studies incorporating body composition and metabolic assessments are needed to clarify the mechanisms underlying these findings and their potential implications for clinical outcomes.

ACKNOWLEDGMENTS

We would like to thank the University Hospital of USP and the entire team of pediatricians and nutritionists from the pediatric ward for their support and assistance.

AUTHORS' CONTRIBUTIONS

DNA: Conceptualization, Data curation, Formal Analysis, Writing – original draft, Writing – review & editing. KGA: Data curation, Formal Analysis, Writing – review & editing. AS: Data curation, Formal Analysis, Writing – review & editing. MTBF: Conceptualization, Data curation, 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. Santiago-Vieira C, Velasquez-Melendez G, Ribeiro-Silva RC, Pinto EJ, Barreto ML, Li L. Recent changes in growth trajectories: a population-based cohort study of over 5 million Brazilian children born between 2001 and 2014. Lancet Reg Health Am. 2024;32:100721. https://doi.org/10.1016/j.lana.2024.100721
2. Fernandes MT, Danti GV, Garcia DM, Ferraro AA. Excess body weight in children may increase the length of hospital stay. Clinics (Sao Paulo). 2015;70(2):87-90. https://doi.org/10.6061/clinics/2015(02)03
3. Domingos IR Jr, Dourado KF, Andrade MIS, Albuquerque MGM, Souza AF, Melo LRS, et al. Prevalence of overweight in children of a referral maternal and child hospital in the state of Pernambuco. Braz J Dev. 2020;6(1):413-29. https://doi.org/10.34117/bjdv6n1-028
4. Kyler KE, Hall M, Bettenhausen JL. Associations between obesity and use for pediatric hospitalizations. Hosp Pediatr. 2020;10(11):997-1000. https://doi.org/10.1542/hpeds.2020-0136
5. Travé TD, García ISM, Benavides AG, Íñigo IV, Aguirre MH, Elízaga AI. Situación nutricional en el momento del ingreso en un hospital pediátrico terciario. Nutr Hosp. 2015;31(6):2465-71. https://doi.org/10.3305/nh.2015.31.6.8863
6. Ramsey KW, Davis J, Okihiro M. A comparison of length of hospitalization and costs in obese and non-obese pediatric patients at a single hospital in Honolulu. Hawaii J Health Soc Welf. 2020;79(5 Suppl 1):91-5. PMID: 32490393
7. Bechard LJ, Rothpletz-Puglia P, Touger-Decker R, Duggan C, Mehta NM. Influence of obesity on clinical outcomes in hospitalized children: a systematic review. JAMA Pediatr. 2013;167(5):476-82. https://doi.org/10.1001/jamapediatrics.2013.13
8. Ferro V, Mosca A, Crea F, Mesturino MA, Olita C, Vania A, et al. The relationship between body mass index and children's presentations to a tertiary pediatric emergency department. Ital J Pediatr. 2018;44:38. https://doi.org/10.1186/s13052-018-0476-y
9. Deng X, Ma J, Yuan Y, Zhang Z, Niu W. Association between overweight or obesity and the risk of childhood asthma and wheeze: an updated meta-analysis of 18 studies including 73,252 children. Pediatr Obes. 2019;14:e12532. https://doi.org/10.1111/ijpo.12532

10. Malden S, Gillespie J, Hughes A, Gibson AM, Farooq A, Martin A, et al. Obesity in young children and its relationship with diagnosis of asthma, vitamin D deficiency, iron deficiency, specific allergies and flat-footedness: a systematic review and meta-analysis. Obes Rev. 2021;22:e13129. https://doi.org/10.1111/obr.13129
11. Aragona E, El-Magbri E, Wang J, Scheckelhoff T, Scheckelhoff T, Hyacinthe A, et al. Impact of obesity on clinical outcomes in urban children hospitalized for status asthmaticus. Hosp Pediatr. 2016;6(4):211-8. https://doi.org/10.1542/hpeds.2015-0094
12. Nyambuya TM, Dludla PV, Mxinwa V, Nkambule BB. Obesity-related asthma in children is characterized by Th1 rather than Th2 immune response: a meta-analysis. Ann Allergy Asthma Immunol. 2020;125(4):425-32. https://doi.org/10.1016/j.anai.2020.06.020
13. Eder K, Baffy N, Falus A, Fulop AK. The major inflammatory mediator interleukin-6 and obesity. Inflamm Res. 2009;58(11):727-36. https://doi.org/10.1007/s00011-009-0060-4
14. Frisancho RA. Anthropometric standards for the assessment of growth and nutritional status. Ann Arbor (MI): University of Michigan Press; 1990.
15. Taylor RW, Jones IE, Williams SM, Goulding A. Evaluation of waist circumference, waist-to-hip ratio, and the conicity index as screening tools for high trunk fat mass, as measured by dual-energy X-ray absorptiometry, in children aged 3-19 years. Am J Clin Nutr. 2000;72(2):490-5. https://doi.org/10.1093/ajcn/72.2.490
16. Onis M, Lobstein T. Defining obesity risk status in the general childhood population: which cut-offs should we use? Int J Pediatr Obes. 2010;5(6):458-60. https://doi.org/10.3109/17477161003615583
17. Ashwell M, Gibson S. A proposal for a primary screening tool: "Keep your waist circumference to less than half your height". BMC Med. 2014;12:207. https://doi.org/10.1186/s12916-014-0207-1
18. Fernandes GXA, Fernandes MLXA, Brito JAA, Carvalho IIA, Sena ASS, Marques TROC. Main causes of hospital admissions of children and adolescents aged 0 to 18 years in the city of Campina Grande, PB, from 2020 to 2022: a cross-sectional study. Rev Foco. 2024;17(2):e4372. https://doi.org/10.54751/revistafoco.v17n2-028
19. Cesare M, D'Agostino F, Nurchis MC, Magliozzi E, Damiani G, Nursing and Public Health Group, et al. Determinants of prolonged hospitalization in children and adolescents: a retrospective observational study. J Nurs Scholarsh. 2025;57(3):412-30. https://doi.org/10.1111/jnu.13045
20. Okubo Y, Michihata N, Yoshida K, Morisaki N, Matsui H, Fushimi K, et al. Impact of pediatric obesity on acute asthma exacerbation in Japan. Pediatr Allergy Immunol. 2017;28(8):763-7. https://doi.org/10.1111/pai.12801
21. Katzow M, Homel P, Rhee K. Factors associated with documentation of obesity in the inpatient setting. Hosp Pediatr. 2017;7(12):731-8. https://doi.org/10.1542/hpeds.2017-0100
22. Tantivit N, Mulekar M, Kaulfers AM, Lim WY. Impact of weight on severity of hospital course in children admitted with COVID-19. Glob Pediatr Health. 2023;10:2333794X231220873. https://doi.org/10.1177/2333794X231220873
23. Wolfe RR. The underappreciated role of muscle in health and disease. Am J Clin Nutr. 2006;84(3):475-82. https://doi.org/10.1093/ajcn/84.3.475
24. Kanneganti TD, Dixit VD. Immunological complications of obesity. Nat Immunol. 2012;13(8):707-12. https://doi.org/10.1038/ni.2343

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260425

Associations of age and sexual abstinence duration with semen parameters in normozoospermic men

Elif Seda Özdemir[1]*, Abuzer Öztürk[2]

SUMMARY

OBJECTIVE: The aim of this study was to evaluate the effects of age and duration of sexual abstinence on semen parameters in normozoospermic men and to explore their relative contributions within physiologically normal ranges.

METHODS: This retrospective study included 412 normozoospermic men defined according to World Health Organization 2010 criteria. Semen volume, sperm concentration, and motility parameters were analyzed. Participants were stratified by age and abstinence duration. Correlation analyses and multivariate linear regression models were performed.

RESULTS: Increasing age was modestly associated with lower total and progressive motility and a higher proportion of immotile sperm. Longer sexual abstinence duration was associated with higher sperm concentration. The proportion of explained variance was modest but consistent across analyses.

CONCLUSION: These findings suggest that age and sexual abstinence duration are associated with selected semen parameters in normozoospermic men. The observed age-related differences in the associations with sexual abstinence and semen parameters warrant further investigation in larger, prospective studies before age-specific abstinence recommendations can be considered.

KEYWORDS: Semen analysis. Sperm motility. Age factors. Sexual abstinence.

INTRODUCTION

Infertility affects approximately 10–20% of couples worldwide, with male factors contributing to nearly half of cases[1]. Semen analysis remains the cornerstone of male fertility evaluation and is standardized according to World Health Organization (WHO) guidelines[2].

Spermatozoa are produced in the seminiferous tubules and mature in the epididymis, where they acquire motility and fertilizing capacity[3,4]. This maturation process, together with ejaculation frequency, contributes to variability in semen parameters[4]. Accordingly, the WHO recommends 2–7 days of sexual abstinence before semen analysis to ensure standardization[2]. Despite this recommendation, the optimal sexual abstinence remains debated[4-7]. While WHO guidelines aim to standardize diagnostic evaluation, reproductive societies often recommend shorter intervals to optimize outcomes in assisted reproduction[8]. In parallel, delayed fatherhood has become increasingly common, raising interest in the effects of paternal age on semen quality. Although age-related changes in semen parameters have been reported, findings remain inconsistent[5]. Moreover, it is unclear whether sexual abstinence has similar effects across different age groups.

Therefore, this study aimed to investigate the associations of age and sexual abstinence with conventional semen parameters in normozoospermic men, with particular focus on whether these associations varied across age groups.

METHODS

Study design and population

This retrospective observational study was conducted at the Andrology Laboratory of Sivas Numune Hospital. Medical records of men who underwent semen analysis between May 2023 and February 2024 were reviewed. Men aged ≥18 years with complete semen data were included. Patients with prior varicocelectomy, grade III varicocele, active fertility treatment, or incomplete data were excluded. Patients with grade III varicocele or a history of varicocelectomy were excluded to reduce potential heterogeneity related to clinically significant varicocele

[1]Sivas Numune Hospital, Andrology Laboratory – Sivas, Turkey.
[2]Sivas Numune Hospital, Department of Urology – Sivas, Turkey.
*Corresponding author: elifsedabilgic@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on August 12, 2026. Accepted on August 15, 2026.
Scientific Editor: José Maria Soares Júnior

and previous surgical intervention. Additionally, individuals with semen parameters below WHO 2010 reference limits were excluded to restrict the cohort to normozoospermic men[2].

Semen collection and analysis

Samples were collected by masturbation after the reported period of sexual abstinence and analyzed according to WHO 2010 guidelines. Sexual abstinence duration was recorded in days according to the laboratory's routine recording system. The predefined abstinence categories were 1–3 and 4–6 days. No participant in the final dataset had a recorded abstinence duration of 1 or 7 days; therefore, the observed abstinence durations ranged from 2 to 6 days.

Statistical analyses

IBM Statistical Package for the Social Sciences (SPSS) Statistics version 20 (Chicago, IL, USA) was used for statistical analyses, and $p<0.05$ was the threshold for statistical significance. The Shapiro-Wilk test was used to examine the normality of the continuous data. Since the data did not exhibit normal distribution, the Mann-Whitney U test was used to compare semen parameters among the age and sexual abstinence groups. Spearman's correlation coefficient was used to examine the relationships between semen parameters, age, and days of sexual abstinence. Multivariate linear regression analysis was used to examine the associations of age and sexual abstinence on semen parameters. Exact two-sided p-values were reported for Mann-Whitney U tests

RESULTS

The distribution of semen parameters according to age group is presented in Table 1. Mean semen volume decreased across age groups, whereas mean sperm concentration increased. Total and progressive motility showed lower mean values with advancing age, while non-progressive and immotile sperm proportions tended to increase.

After adjustment for sexual abstinence duration, each 1-year increase in age was associated with a 0.024 mL decrease in semen volume across the study population. In contrast, sperm concentration and the proportion of immotile sperm increased by 0.703×10^6/mL and 0.203%, respectively, per year of age. Total motility and progressive motility decreased by 0.191 and 0.240%, respectively, with each additional year of age. When adjusting for age, each additional day of sexual abstinence was associated with an increase of 5.929×10^6/mL in sperm concentration (Table 2).

In age-stratified analyses, a significant difference in semen volume was observed among men aged 20–29 years according to sexual abstinence. Semen volume was higher in individuals who abstained for 4–6 days compared with those who abstained for 1–3 days (p=0.01). Among men aged 30–39 years, sperm concentration was significantly higher in those with an abstinence period of 4–6 days compared with shorter sexual abstinence (p=0.03) (Table 3).

Correlation analyses demonstrated negative linear correlations between age and semen volume, total motility, and progressive motility, as well as a positive linear correlation between age and the proportion of immotile sperm (Figure 1). In addition, a positive linear correlation was observed between sexual abstinence and sperm concentration, indicating higher sperm concentrations with longer periods of abstinence (Table 4).

DISCUSSION

Male reproductive function undergoes gradual changes with advancing age. Although male fertility can be preserved throughout life, age-related alterations in androgen production, spermatogenesis, and sexual function have been well documented. Increasing evidence suggests that advanced paternal age is associated with changes in semen quality, fertility potential, pregnancy outcomes, and offspring health[9]. Because age-related differences in sexual activity may also contribute

Table 1. Distribution of semen parameters according to age group.

Semen parameter	20–29 years (n=228) mean (95%CI)	30–39 years (n=157) mean (95%CI)	40–49 years (n=27) mean (95%CI)
Semen volume (mL)	3.5 (3.3–3.7)	3.2 (3.1–3.5)	2.9 (2.5–3.3)
Sperm concentration (million/mL)	64.7 (59.9–69.5)	69.6 (62.6–76.6)	85.4 (66.2–104.6)
Total motility (%)	67.0 (65.5–68.5)	66.3 (64.6–68.0)	61.5 (57.0–66.0)
Progressive motility (%)	55.7 (54.1–57.3)	54.1 (52.1–56.0)	48.7 (44.3–53.1)
Non-progressive motility (%)	11.4 (10.8–12.1)	12.4 (11.6–13.3)	12.9 (10.6–15.1)
Immotile sperm (%)	33.1 (31.6–34.7)	34.0 (32.3–35.7)	38.8 (34.3–43.3)

CI: confidence interval.

Table 2. Multivariate linear regression analysis of age and sexual abstinence duration in relation to semen parameters.

		B	SE	β	95%CI	p	Model R^2
Semen volume (mL)	Age	-0.024	0.011	-0.109	-0.046 to -0.003	**0.027**	0.018
	Sexual abstinence	0.151	0.089	0.083	-0.024 to 0.325	0.091	
Sperm concentration (10^6/mL)	Age	0.703	0.327	0.105	0.060 to 1.347	**0.032**	0.024
	Sexual abstinence	5.929	2.640	0.110	0.739 to 11.120	**0.025**	
Total motility (%)	Age	-0.191	0.091	-0.103	-0.369 to -0.012	**0.036**	0.013
	Sexual abstinence	-0.605	0.732	-0.041	-2.043 to 0.833	0.409	
Progressive motility (%)	Age	-0.238	0.099	-0.118	-0.433 to -0.043	**0.017**	0.018
	Sexual abstinence	-0.959	0.803	-0.059	-2.532 to 0.626	0.233	
Non-progressive motility (%)	Age	0.053	0.042	0.062	-0.030 to 0.136	0.210	0.007
	Sexual abstinence	0.384	0.341	0.056	-0.285 to 1.054	0.260	
Immotile sperm (%)	Age	0.203	0.091	0.109	0.024 to 0.382	**0.027**	0.013
	Sexual abstinence	0.532	0.736	0.035	-0.915 to 1.978	0.471	

B: unstandardized regression coefficient; SE: standard error; β: standardized regression coefficient; 95%CI: 95% confidence interval; R^2: coefficient of determination. Bold values indicate statistical significance ($p<0.05$).

Table 3. Comparison of semen parameters according to sexual abstinence duration within age groups.

Age group	Semen parameter	1–3 days, mean (95%CI)	4–6 days, mean (95%CI)	Mann-Whitney U	Exact p-value
		n=120	n=108		
20–29 years (n=228)	Semen volume (mL)	3.32 (3.07–3.56)	3.79 (3.50–4.08)	5,191.0	0.009
	Sperm concentration (×10^6/mL)	62.39 (55.52–69.26)	67.25 (60.50–74.01)	5,878.0	0.227
	Total motility (%)	66.89 (64.81–68.97)	67.07 (64.89–69.26)	6,399.0	0.871
	Progressive motility (%)	55.97 (53.72–58.21)	55.34 (53.04–57.64)	6,312.5	0.737
	Non-progressive motility (%)	11.02 (10.03–12.00)	11.90 (10.96–12.84)	5,909.5	0.251
	Immotile sperm (%)	33.38 (31.28–35.48)	32.87 (30.63–35.11)	6,302.5	0.722
30–39 years (n=157)		n=70	n=87		
	Semen volume (mL)	3.34 (3.02–3.67)	3.26 (3.00–3.53)	2,968.5	0.785
	Sperm concentration (×10^6/mL)	60.32 (51.44–69.20)	77.05 (66.74–87.36)	2,438.5	0.032
	Total motility (%)	67.47 (64.80–70.15)	65.36 (63.13–67.59)	2,712.5	0.241
	Progressive motility (%)	55.41 (52.22–58.61)	52.97 (50.43–55.50)	2,730.5	0.268
	Non-progressive motility (%)	12.23 (10.85–13.61)	12.62 (11.65–13.60)	2,901.5	0.613
	Immotile sperm (%)	32.83 (30.18–35.48)	35.00 (32.79–37.21)	2,713.0	0.242
40–49 years (n=27) (exploratory)		n=14	n=13		
	Semen volume (mL)	3.07 (2.48–3.66)	2.69 (2.11–3.28)	70.5	0.312
	Sperm concentration (×10^6/mL)	84.64 (61.44–107.84)	86.31 (51.31–121.31)	90.5	0.991
	Total motility (%)	62.50 (55.30–69.70)	60.38 (53.99–66.78)	82.0	0.676
	Progressive motility (%)	48.50 (41.55–55.45)	48.92 (42.66–55.19)	90.0	0.971
	Non-progressive motility (%)	14.21 (10.53–17.90)	11.46 (8.63–14.29)	63.5	0.187
	Immotile sperm (%)	37.79 (30.57–45.00)	39.85 (33.54–46.15)	83.0	0.711

Data are presented as mean (95%CI). P-values were calculated using the Mann-Whitney U test with exact two-sided significance values. Bold values indicate statistical significance ($p<0.05$). CI: confidence interval.

to reproductive outcomes[10], sexual abstinence represents a potentially relevant behavioral factor when interpreting age-related differences in semen parameters. The optimal duration of sexual abstinence remains controversial. Current guidelines differ in their recommendations, with WHO suggesting 2–7 days for standard semen analysis, whereas European Society of Human Reproduction and Embryology (ESHRE) recommends shorter intervals (3–4 days) in the assisted reproduction[2,8].

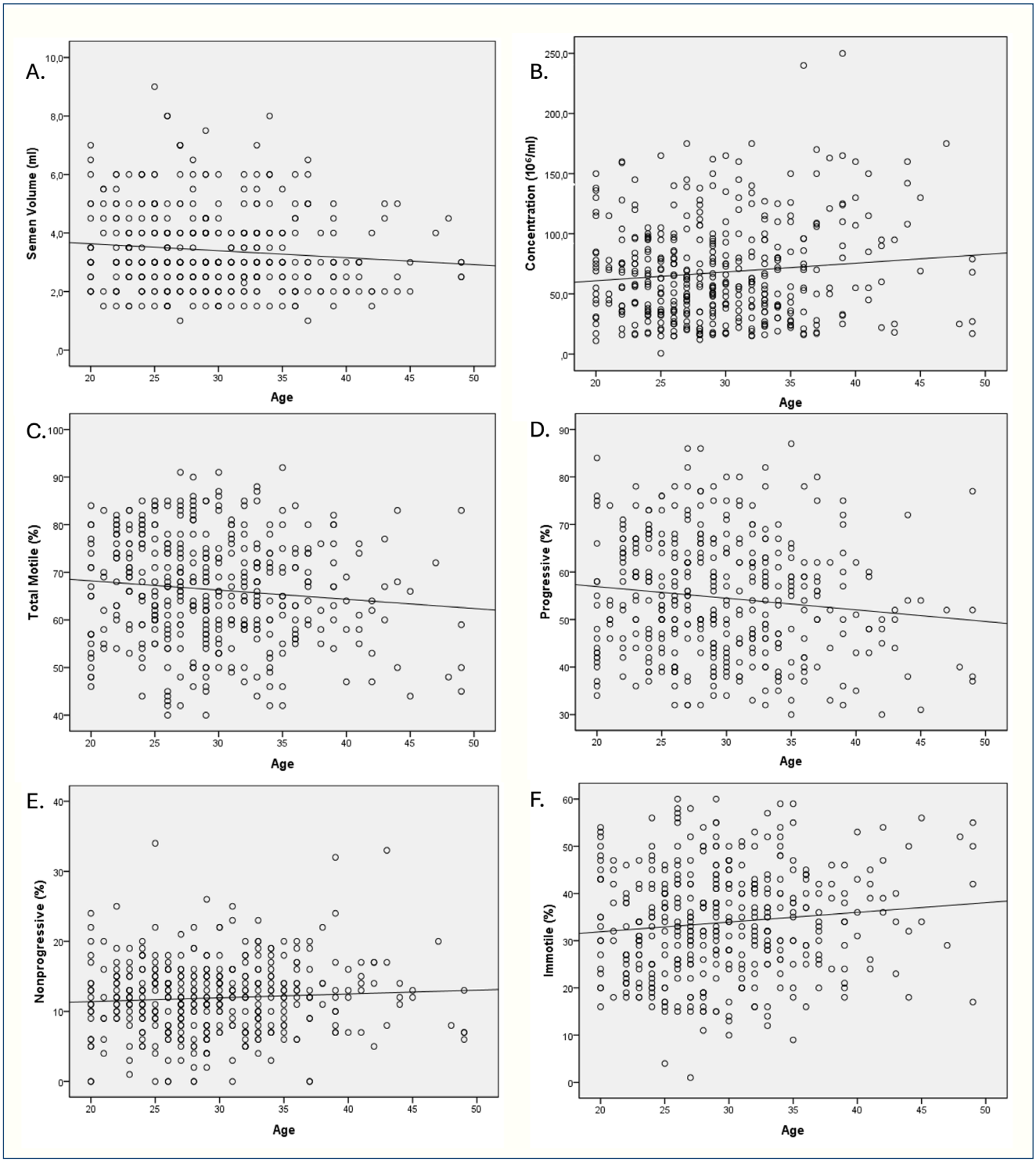


Figure 1. Scatterplots illustrating the unadjusted relationships between age and semen parameters. (A) Semen volume, (B) sperm concentration, (C) total motility, (D) progressive motility, (E) non-progressive motility, and (F) immotile sperm.

Table 4. Spearman correlations between age, sexual abstinence duration, and semen parameters.

	Age		Sexual abstinence	
	r_s	p	r_s	p
Semen volume (mL)	-0.097	**0.04**	0.086	0.08
Concentration (10^6/mL)	0.052	0.29	0.109	**0.02**
Total motile (%)	-0.101	**0.04**	-0.041	0.40
-Progressive motility (%)	-0.113	**0.02**	-0.059	0.23
-Non-progressive motility (%)	0.049	0.31	0.049	0.32
Immotile sperm (%)	0.105	**0.03**	0.034	0.48

r_s: Spearman's rank correlation coefficient; p: two-sided p-value. Bold values indicate statistical significance ($p<0.05$).

These recommendations are largely based on different clinical endpoints and therefore may not be directly interchangeable. Previous studies have reported differences in reproductive outcomes according to abstinence duration, including possible associations between shorter abstinence intervals and fertilization or pregnancy outcomes in assisted reproduction[11]. However, these findings involve different clinical endpoints from those evaluated in routine semen analysis. Therefore, the present study sought to explore whether the association between sexual abstinence and conventional semen parameters varied across age groups within a normozoospermic population.

In the present study, we evaluated semen volume, concentration, and motility in a strictly normozoospermic population, focusing on the associations of age and sexual abstinence with conventional semen parameters. We observed a modest decline in semen volume with increasing age, consistent with previous reports[12,13]. This finding has been attributed to age-related functional changes in the seminal vesicles, which contribute substantially to ejaculate volume and fructose secretion[5]. However, conflicting evidence exist, reporting either no association[14] or even an increase in semen volume with age[15], suggesting potential confounding factors.

Longer abstinence periods were associated with higher semen volume in men in their twenties. This observation may reflect enhanced secretory activity of the prostate and seminal vesicles following androgen stimulation[16]. Age-related differences in androgen-dependent accessory gland function may modulate the effect of sexual abstinence on semen volume, potentially making this relationship more apparent in younger men. Further studies are required to confirm this age-dependent effect.

In age-stratified analyses, longer abstinence was significantly associated with higher semen volume among men aged 20–29 years, whereas among men aged 30–39 years, longer abstinence was significantly associated with higher sperm concentration. No significant associations were observed for the other semen parameters within these age groups. The age-stratified findings provide preliminary evidence of potential age-dependent associations between sexual abstinence and semen parameters. Further studies with larger and more balanced age-stratified populations are needed to confirm the reproducibility and clinical relevance of these findings.

Sperm concentration increased with advancing age in our cohort, consistent with some previous studies[12,17], although contradictory findings have also been reported[14,18]. Given the concurrent decline in semen volume with age, the observed increase in sperm concentration may partly reflect a concentration effect rather than increased spermatogenic output. Regarding sexual abstinence, we observed a positive association between longer abstinence and higher sperm concentration, consistent with previous evidence demonstrating a positive association between abstinence duration and sperm concentration[19,20]. Recent large-scale evidence has similarly demonstrated an increase in sperm concentration with longer abstinence among normozoospermic men, while no significant association with motility was observed[21].

The selection of a normozoospermic population was intentional, as it allowed the associations between age, sexual abstinence, and conventional semen parameters to be evaluated within a more homogeneous semen profile. However, this design may have limited the range of semen parameter values and consequently reduced the ability to detect stronger associations. Recent evidence further supports the relevance of studying age-related changes even among normozoospermic men. In a 2026 study of 534 normozoospermic men, advancing paternal age was associated with lower ejaculatory volume and progressive motility, as well as higher sperm DNA fragmentation, indicating that age-related alterations may be detectable even when conventional semen parameters remain within the normal range[22]. These findings support our rationale for examining age-related associations within a normozoospermic population rather than restricting the analysis to men with overt semen abnormalities.

In the multivariable analysis, each additional day of sexual abstinence was associated with an increase in sperm concentration, although the relatively wide 95%CI indicates uncertainty around the magnitude of this association. This finding should therefore be interpreted as a modest population-level association rather than as a basis for predicting an individual patient's sperm concentration.

Total and progressive sperm motility declined with increasing age, in agreement with the majority of published data[6]. These changes may be related to age-associated dysfunction of post-testicular organs, particularly the epididymis, where sperm acquire motility[23]. Age-related reductions in prostate- and epididymis-derived secretory markers, as well as increased oxidative stress, may contribute to impaired sperm motility[24].

The relationship between sexual abstinence and sperm motility remains controversial. While several studies report reduced motility with prolonged abstinence, these findings often involve oligospermic or asthenozoospermic populations[25]. A recent study of 3,052 men undergoing fertility evaluation similarly found that longer abstinence was associated with higher sperm concentration but lower progressive and total motility[26]. In our normozoospermic cohort, sexual abstinence duration was not significantly associated with total, progressive, non-progressive, or immotile sperm proportions in the overall multivariable analysis. Furthermore, the age-stratified analyses did not demonstrate significant differences in motility-related parameters according to abstinence duration.

In accordance with the study protocol, individuals with semen parameters below the WHO 2010 lower reference limits (<5th percentile) were excluded, resulting in a cohort restricted to normozoospermic men. While this approach improved cohort homogeneity, it may also have introduced range restriction by reducing variability in semen parameters and potentially attenuating the magnitude of detectable associations. Accordingly, the present findings should be interpreted within the context of a normozoospermic population and should not be directly generalized to men with abnormal semen parameters.

In addition, several limitations should be considered. Sperm morphology was not included, because it was not recorded in a sufficiently standardized and consistently available manner in the retrospective dataset; therefore, a fully comprehensive assessment of semen quality was not possible. The relatively small proportion of older individuals, particularly the 40–49-year age group (n=27), may limit the precision and generalizability of the age-stratified findings, and results for this subgroup should therefore be considered exploratory. Furthermore, the retrospective design limited our ability to adjust for potential confounding factors, including BMI, smoking, alcohol consumption, comorbidities, and medication use, and residual confounding cannot be excluded. The single-center design may also limit the generalizability of the findings to other populations and geographic settings, as semen parameters may vary across populations. The relatively low R^2 values across the regression models represent an additional limitation, as age and sexual abstinence duration accounted for only a limited proportion of the variability in the examined semen parameters. This finding is consistent with the multifactorial nature of semen characteristics and the limited number of covariates available in our retrospective dataset. Accordingly, the statistically significant associations observed should be interpreted as modest population-level associations rather than individually predictive effects or as a basis for clinical thresholds or age-specific abstinence recommendations.

At the same time, the study has several strengths, including the use of standardized WHO 2010 criteria, single-laboratory analysis, and multivariate modeling, allowing the associations of age and sexual abstinence to be evaluated simultaneously. Age was associated with several semen parameters, particularly motility-related parameters, whereas sexual abstinence showed its most consistent association with sperm concentration.

Our findings suggest that the associations between sexual abstinence duration and conventional semen parameters may vary across age groups in normozoospermic men. These age-dependent findings should be considered exploratory and hypothesis-generating rather than as a basis for age-specific abstinence recommendations, and warrant confirmation in larger prospective and multicenter studies.

CONCLUSION

Advancing age was associated with measurable changes in semen quality in normozoospermic men. Longer sexual abstinence duration was associated with higher sperm concentration, while age-stratified analyses suggested that the relationship between abstinence duration and selected semen parameters may vary across age groups. These findings highlight the importance of considering age when interpreting semen analysis and warrant further investigation in larger, prospective studies to determine whether age-related differences in sexual abstinence duration have clinically meaningful implications.

AUTHOR'S NOTE

The authors utilized OpenAI's ChatGPT to support improvements in grammar, language clarity, and manuscript phrasing. All study design, data analysis, interpretation of results, and scientific content were developed and verified by the authors.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study was approved by the ethics committee of the Cumhuriyet University (Ref. No.: 2024-02/26), and conducted in accordance with the Helsinki Declaration.

AUTHORS' CONTRIBUTIONS

ESÖ: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review & editing. **AÖ:** Conceptualization, Data curation, Formal Analysis, 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. Onofre J, Geenen L, Cox A, Auwera I, Willendrup F, Andersen E, et al. Simplified sperm testing devices: a possible tool to overcome lack of accessibility and inconsistency in male factor infertility diagnosis. An opportunity for low-and middle-income countries. Facts Views Vis ObGyn. 2021;13(1):79.
2. WHO. Laboratory manual for the examination and processing of human semen. 5th ed. Geneva: World Health Organisation Press; 2010.
3. Sullivan R, Mieusset R. The human epididymis: its function in sperm maturation. Hum Reprod Update. 2016;22(5):574-87.
4. Mayorga-Torres BJM, Camargo M, Agarwal A, Plessis SS, Cadavid ÁP, Cardona Maya WD. Influence of ejaculation frequency on seminal parameters. Reprod Biol Endocrinol. 2015;13:1-7.
5. Kidd SA, Eskenazi B, Wyrobek AJ. Effects of male age on semen quality and fertility: a review of the literature. Fertil Steril. 2001;75(2):237-48.
6. Johnson SL, Dunleavy J, Gemmell NJ, Nakagawa S. Consistent age-dependent declines in human semen quality: a systematic review and meta-analysis. Ageing Res Rev. 2015;19:22-33.
7. Du C, Li Y, Yin C, Luo X, Pan X. Association of abstinence time with semen quality and fertility outcomes: a systematic review and dose–response meta-analysis. Andrology. 2024;12(6):1224-35.
8. Björndahl L, Kvist U. Manual on basic semen analysis. ESHRE Monographs 2. Oxford: Oxford University Press; 2002.
9. Kühnert B, Nieschlag E. Reproductive functions of the ageing male. Hum Reprod Update. 2004;10(4):327-39.
10. Sharma R, Agarwal A, Rohra VK, Assidi M, Abu-Elmagd M, Turki RF. Effects of increased paternal age on sperm quality, reproductive outcome and associated epigenetic risks to offspring. Reprod Biol Endocrinol. 2015;13(1):35.
11. Hanson BM, Aston KI, Jenkins TG, Carrell DT, Hotaling JM. The impact of ejaculatory abstinence on semen analysis parameters: a systematic review. J Assisted Reprod Genet. 2018;35(2):213-20.
12. Kovac JR, Addai J, Smith RP, Coward RM, Lamb DJ, Lipshultz LI. The effects of advanced paternal age on fertility. Asian J Androl. 2013;15(6):723.
13. Pino V, Sanz A, Valdés N, Crosby J, Mackenna A. The effects of aging on semen parameters and sperm DNA fragmentation. JBRA Assisted Reprod. 2020;24(1):82.
14. Blachman-Braun R, Best JC, Sandoval V, Lokeshwar SD, Patel P, Kohn T, et al. Sperm DNA fragmentation index and high DNA stainability do not influence pregnancy success after intracytoplasmic sperm injection. F&S Rep. 2020;1(3):233-8.
15. Carlsen E, Swan SH, Petersen JH, Skakkebæk NE. Longitudinal changes in semen parameters in young Danish men from the Copenhagen area. Hum Reprod. 2005;20(4):942-9.
16. Cheng H, Zhang X, Li Y, Cao D, Luo C, Zhang Q, et al. Age-related testosterone decline: mechanisms and intervention strategies. Reprod Biol Endocrinol. 2024;22(1):144.
17. Brahem S, Mehdi M, Elghezal H, Saad A. The effects of male aging on semen quality, sperm DNA fragmentation and chromosomal abnormalities in an infertile population. J Assisted Reprod Genet. 2011;28:425-32.
18. Guo LY, Zhou H, Liu M, Li Q, Sun XF. Male age is more critical to sperm DNA integrity than routine semen parameters in Chinese infertile males. Andrologia. 2020;52(1):e13449.
19. Alipour H, Duus RK, Wimmer R, Dardmeh F, Plessis S, Jørgensen N, et al. Seminal plasma metabolomics profiles following long (4–7 days) and short (2 h) sexual abstinence periods. Eur J Obstetr Gynecol Reprod Biol. 2021;264:178-83.
20. Giudice A, Asmundo MG, Cimino S, Cocci A, Falcone M, Capece M, et al. Effects of long and short ejaculatory abstinence on sperm parameters: a meta-analysis of randomized-controlled trials. Front Endocrinol (Lausanne). 2024;15:1373426.
21. Russo GI, Asmundo MG, Cocci A, Abdelhameed AS, Liprino A, Giacone F, et al. Ejaculatory abstinence duration impacts semen parameters: insights from a retrospective analysis in male infertility on 23,527 analyses. Front Endocrinol (Lausanne). 2025;16:1529262.
22. Liu CM, Chen HY, Hou HC, Kuo JH, Lin TY, Lin YM, et al. The reduction in total motile sperm count and the increase in sperm DNA fragmentation are associated with paternal age in normozoospermic men. Aging Male. 2026;29(1):2612794.
23. Gagnon C, Lamirande E. Controls of sperm motility. The sperm cell: production, maturation, fertilization, regeneration. Cambridge; 2006. p. 108-33.
24. Aitken RJ, Drevet JR. The importance of oxidative stress in determining the functionality of mammalian spermatozoa: a two-edged sword. Antioxidants. 2020;9(2):111.
25. Kulkarni V, Kaingade P, Kulkarni N, Bhalerao T, Nikam A. Assessment of semen parameters in consecutive ejaculates with short abstinence period in oligospermic males. JBRA Assisted Reprod. 2022;26(2):310.
26. Yao G, Qi Q, Dou X, Zhou W, Bai S, Zhang X. Association of abstinence time with semen quality in men who undergoing fertility evaluation: a cross-sectional study from 3052 participants. Front Endocrinol (Lausanne). 2025;16:1472333.

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260486

Female sexual dysfunction in anti-synthetase syndrome

Lorenza Rosa Silvério da Silva[1]*, Theo Lerner[2], Elsa Aida Gay de Pereyra[2], Alexandre Moura dos Santos[1], Samuel Katsuyuki Shinjo[1]

SUMMARY

OBJECTIVE: Sexual dysfunction has not been investigated in women with anti-synthetase syndrome. The aim of this study was to determine the prevalence of sexual dysfunction in women with anti-synthetase syndrome compared with healthy controls (CTR), to evaluate associated clinical factors, and to assess the effect of cognitive behavioral therapy on sexual function and related outcomes.

METHODS: This single-center study included a cross-sectional evaluation conducted between March and August 2023 and a prospective interventional phase. Thirty-three women with anti-synthetase syndrome and 31 CTR were included. Sexual function, sexual distress, quality of life, fatigue, depression, body image, physical activity, and disease-related parameters were assessed using validated questionnaires. Multivariate logistic regression was used to identify factors independently associated with sexual dysfunction. Patients with anti-synthetase syndrome and sexual dysfunction were invited to participate in group-based cognitive behavioral therapy, with up to 90 days' follow-up.

RESULTS: Groups were comparable with respect to age and marital status. Sexual dysfunction prevalence was significantly higher in the anti-synthetase syndrome group than in CTR (81.8 vs. 51.6%; p=0.016). Anti-synthetase syndrome was independently associated with sexual dysfunction (OR 4.11; 95%CI 1.20–15.78). Within the anti-synthetase syndrome cohort, patients with and without sexual dysfunction were similar across most variables, except for prednisone use, which occurred only in patients with sexual dysfunction. Six patients completed cognitive behavioral therapy, showing improvements in sexual distress, sexual function—particularly arousal—and fatigue at 90 days.

CONCLUSION: Sexual dysfunction is highly prevalent in women with anti-synthetase syndrome and appears intrinsically associated with the disease. Cognitive behavioral therapy represents a promising non-pharmacological strategy to improve sexual function, distress, and fatigue in this population.

KEYWORDS: Antisynthetase syndrome. Idiopathic inflammatory myopathies. Sexual dysfunction. Physiological. Women. Cognitive behavioral therapy.

INTRODUCTION

Anti-synthetase syndrome (ASyS) is a distinct subset of idiopathic inflammatory myopathies characterized by the presence of anti-aminoacyl-transfer-ribonucleic acid synthetase autoantibodies and a heterogeneous, multisystem clinical spectrum. Classically, ASyS is defined by the triad of inflammatory myopathy, arthritis, and interstitial lung disease; however, disease expression is highly variable and may evolve throughout the disease course. Additional manifestations such as fever, Raynaud's phenomenon, and hyperkeratotic fissuring of the fingers, commonly referred to as "mechanic's hands," are frequently observed and may contribute to functional impairment, occupational limitations, and reduced quality of life. Anti-Jo-1 is the most prevalent anti-synthetase antibody, although several other autoantibodies have been described and may be associated with distinct clinical phenotypes and prognostic profiles[1].

Sexual health is a fundamental component of global health and well-being, encompassing biological, psychological, interpersonal, and sociocultural dimensions. It extends beyond sexual activity itself and includes sexual identity, intimacy, emotional connection, self-esteem, and satisfaction. Sexual dysfunction (SD) is defined as persistent or recurrent disturbance in sexual desire, arousal, orgasm, or satisfaction that results in clinically significant distress or interpersonal difficulty[2-4]. These disturbances arise from complex interactions involving neuroendocrine regulation, vascular integrity, musculoskeletal function, emotional processing, cognitive factors, and contextual influences such as relationship dynamics and cultural background.

Systemic autoimmune rheumatic diseases (SARDs) are consistently linked with impaired sexual health. Chronic pain, fatigue, stiffness, physical disability, mucocutaneous dryness, altered body image, depressive and anxiety symptoms, hormonal dysregulation, and adverse effects of pharmacological

[1]Universidade de São Paulo, Faculdade de Medicina, Division of Rheumatology – São Paulo (SP), Brazil.
[2]Universidade de São Paulo, Faculdade de Medicina, Hospital das Clínicas, Division of Obstetrics and Gynecology – São Paulo (SP), Brazil.
*Corresponding author: lorenzasilverio@gmail.com

Conflicts of interest: the authors declare there is no conflicts of interest. Funding: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) [#2020/10691-4] to AMS; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [#301500/2022-3], and Faculdade de Medicina da USP to SKS.
Received on April 25, 2026. Accepted on July 29, 2026.
Scientific Editor: Roseli Nomura

therapies may all contribute to SD in this population[5-7]. In addition, uncertainty regarding disease prognosis, fear of symptom exacerbation during sexual activity, and changes in partner dynamics may further impair sexual well-being. Despite its high prevalence and substantial impact on quality of life, sexual health remains underrecognized and underdressed in routine rheumatologic care. Many patients report embarrassment, lack of opportunity, or discomfort in discussing sexual concerns with healthcare professionals, resulting in unmet needs and persistent distress[7].

It is well established that multiple factors contribute to SD—such as medical conditions, medication use, psychological factors, and relationship-specific issues[8]. However, although it is known that these factors lead to reduced sexual activity, there are no specific pathophysiological mechanisms that fully explain the phenomenon in isolation; this supports the biopsychosocial hypothesis of female sexual function, which integrates neuroendocrinological, psychosocial, and personal factors.

Within idiopathic inflammatory myopathies, SD has been documented in both men and women and has been associated with muscle weakness, reduced physical activity, fatigue, depressive symptoms, and impaired quality of life[9]. Importantly, these associations appear to be largely independent of disease duration or type of immunosuppressive therapy, suggesting that SD may represent a core disease-related burden rather than a secondary consequence of treatment or chronicity. In ASyS specifically, interstitial lung disease and its associated dyspnea and reduced exercise tolerance, chronic glucocorticoid exposure, persistent fatigue, disease-related changes in body image, psychological distress, and systemic inflammatory activity have been proposed as plausible, disease-specific contributors to impaired sexual function, consistent with the biopsychosocial model of female sexual health endorsed by the International Society for the Study of Women's Sexual Health and the International Consultation on Sexual Medicine[3,8]. However, to date, no study has specifically evaluated sexual function in patients with ASyS, a subgroup characterized by systemic involvement, pulmonary disease, long-term immunosuppression, and complex psychosocial challenges.

Given the absence of disease-specific data and the growing recognition of sexual health as an essential domain of patient-centered care, this study aimed to: (1) determine the prevalence of SD in adult female patients with ASyS compared with healthy controls; (2) identify demographic, clinical, laboratory, and psychosocial factors associated with SD; and (3) prospectively assess the impact of cognitive-behavioral therapy (CBT) on sexual function, sexual distress, and related outcomes in women with ASyS.

METHODS

Study design and participants

This single-center study consisted of a cross-sectional observational phase followed by a prospective interventional phase. Adult female patients with ASyS receiving regular follow-up at a tertiary rheumatology center were consecutively invited to participate between March and August 2023. ASyS was defined according to the criteria proposed by Behrens Pinto et al.[10]. Only patients with stable disease and regular outpatient follow-up were considered eligible to minimize confounding related to acute disease exacerbations.

Exclusion criteria included pregnancy; active disease requiring escalation of immunosuppressive therapy, high-dose glucocorticoids, intravenous immunoglobulin, or pulse therapy; overlap syndromes with other connective tissue diseases; history of gynecological malignancy or pelvic irradiation; and active gynecological infection at the time of evaluation. These criteria were applied to minimize potential bias related to conditions independently associated with SD.

A control group was composed of adult women without rheumatologic diseases who completed electronic questionnaires during the same period. Controls were recruited to provide a reference population for sexual function outcomes and were required to meet the same exclusion criteria related to gynecological conditions.

Data collection and assessments

All participants completed standardized questionnaires addressing sociodemographic characteristics, educational level, socioeconomic status, marital and relationship status, gynecological history, comorbidities, lifestyle habits, and anthropometric data. Sexual function was assessed using validated Brazilian Portuguese versions of the Female Sexual Function Index (FSFI)[11], the Female Sexual Distress Scale–Revised (FSDS-R)[12], and the Female Sexual Quotient (FSQ)[13]. SD was defined as FSFI ≤26.55 and/or FSQ ≤60, while sexual distress was defined as FSDS-R ≥11.

Disease activity and functional status were evaluated using Manual Muscle Testing (MMT-8), Myositis Disease Activity Assessment Visual Analog Scales (MYOACT), and the Health Assessment Questionnaire (HAQ). Physician- and patient-reported global assessments were obtained using visual analog scales to capture both clinical and subjective perceptions of disease impact. Fatigue was assessed using the Fatigue Severity Scale, depressive symptoms using the Beck Depression Inventory, body image dissatisfaction using the Body Shape Questionnaire, physical activity using the International Physical

Activity Questionnaire–Short Form, pain using a visual analog scale, and quality of life using the EQ-5D questionnaire[14-16].

Laboratory parameters, including serum muscle enzymes and myositis-specific and myositis-associated autoantibody profiles, as well as current and prior pharmacological treatments, were obtained from medical records and reviewed by a single investigator to ensure data consistency[17].

Prospective intervention

Patients with ASyS who met criteria for SD were invited to participate in a group-based CBT intervention. Sessions were conducted weekly over a 12-week period and followed a previously validated semi-structured protocol[18]. The intervention focused on increasing sexual self-awareness, addressing dysfunctional beliefs related to sexuality, reducing performance anxiety, improving communication skills, and promoting adaptive coping strategies. Sexual function questionnaires were administered before the intervention and at 30 and 90 days after completion to assess short- and medium-term effects.

Statistical analysis

Continuous variables were expressed as means±standard deviation or medians with interquartile ranges, as appropriate. Categorical variables were presented as absolute and relative frequencies. The normality of continuous variables was assessed using the Shapiro-Wilk test. Normally distributed continuous variables were compared between groups using Student's t-test, whereas non-normally distributed variables were compared using the Mann-Whitney U test; categorical variables were compared using the chi-square test or Fisher's exact test, as appropriate. Group comparisons were performed using parametric or non-parametric tests according to data distribution. Multivariate logistic regression analysis was applied to identify factors independently associated with SD. Candidate variables for inclusion in the multivariate model were selected based on their statistically significant association with SD in univariate analysis ($p<0.05$); the study group (ASyS vs. controls) was retained in the model as the primary variable of interest, given the central hypothesis of the study. Missing data were not imputed; participants with missing values for a given variable were excluded from the analyses involving that specific variable (complete-case analysis). Statistical significance was set at $p<0.05$. Statistical power analysis and computational procedures followed established methodological standards[19,20]. Given the limited sample size, particularly in the prospective interventional phase, inferential results from that phase should be interpreted with caution and regarded as exploratory and hypothesis-generating rather than confirmatory.

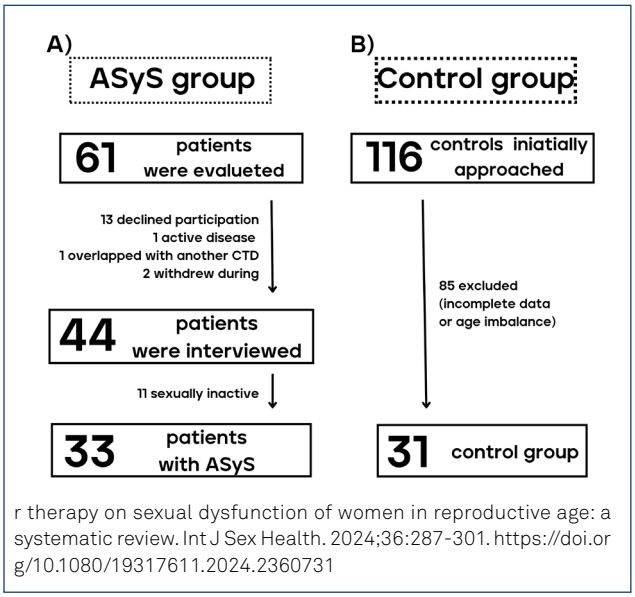
RESULTS

Cross-sectional analysis

A total of 33 women with ASyS and 31 healthy controls were included in the analysis (Figure 1). Groups were comparable regarding age and marital status. Patients with ASyS presented lower educational attainment, poorer socioeconomic status, and higher body mass index compared with controls.

The prevalence of SD was significantly higher in the ASyS group than in controls (81.8 vs. 51.6%; p=0.016). In multivariate analysis, ASyS was independently associated with SD (OR 4.11; 95%CI 1.20–15.78). Candidate variables entered into the multivariate model were median age, race, economic status, educational level, and body mass index, all of which showed a statistically significant association with SD in univariate analysis ($p<0.05$); no candidate variable was excluded from the model. After adjustment, ASyS remained independently and significantly associated with SD, whereas none of the other covariates retained a significant association with SD (OR>1 for all; $p>0.50$ for all).

Within the ASyS cohort, patients with and without SD did not differ significantly in terms of disease activity scores, laboratory parameters, autoantibody profiles, fatigue, depressive symptoms, pain, body image dissatisfaction, quality of life, or physical activity. Prednisone use was observed exclusively among patients with SD.



r therapy on sexual dysfunction of women in reproductive age: a systematic review. Int J Sex Health. 2024;36:287-301. https://doi.org/10.1080/19317611.2024.2360731

Figure 1. Flowchart of participant selection. (A) Selection of participants with anti-synthetase syndrome. (B) Selection of participants in the control group. ASyS: anti-synthetase syndrome; CTD: connective tissue disease.

Prospective analysis

All ASyS patients with SD were invited to participate in CBT. Eight patients initially enrolled, and six completed the intervention (Table 1). At 90-day follow-up, significant improvements were observed in sexual distress and sexual function scores, with particular improvement in the arousal domain. Fatigue scores also decreased after the intervention, while disease activity and other psychosocial measures remained unchanged.

DISCUSSION

In this study, SD was highly prevalent among women with ASyS, exceeding that observed in healthy controls and aligning with rates reported in other SARDs[2,5,7]. Importantly, SD was independently associated with the diagnosis of ASyS, rather than with disease activity, laboratory markers, or psychosocial variables, suggesting that SD may represent an intrinsic and underrecognized burden of the disease. The prevalence of SD in the control group (51.6%) aligns with the global prevalence of 50.8% reported by Alidost et al[21].

The absence of associations between SD and fatigue, depression, or body image contrasts with findings reported in other rheumatic diseases and may reflect unique pathophysiological mechanisms or psychosocial dynamics specific to ASyS. Chronic pulmonary involvement, fluctuating disease course, and long-term exposure to glucocorticoids may contribute to altered sexual function through complex biological and psychological pathways. Although prednisone use was observed exclusively among patients with SD, causal inference is limited by the cross-sectional design. Further investigation is warranted, and future multicenter, longitudinal studies may help clarify this issue.

Prospective findings suggest a potential benefit of CBT, which may be a potentially effective non-pharmacological intervention for improving sexual function and reducing sexual distress in women with ASyS. These findings are considered exploratory and hypothesis-generating, requiring validation in larger controlled prospective studies. Improvements were observed particularly in domains related to sexual arousal and distress, consistent with previous studies demonstrating the efficacy of CBT and mindfulness-based

Table 1. General characteristics of patients with anti-synthetase syndrome and sexual dysfunction with behavioral therapy intervention.

Parameters	Baseline	D90	p-value
Disease status			
HAQ (0.00–3.00)	0.13 (0.00–0.72)	0.50 (0.00–1.00)	0.789
MMT-8 (0–80)	80 (78–80)	78 (76–78)	0.169
MYOACT	0.0 (0.0–0.2)	0.1 (0.0–0.2)	0.427
PGA (0–10)	6.0 (1.3–7.0)	6.0 (1.5–6.0)	0.587
PhGA (0–10)	0.0 (0.0–3.8)	0.0 (0.0–2.4)	0.346
Laboratory data			
CPK (U/L)	182 (80–396)	258 (137–398)	0.562
LDH (U/L)	234 (198–239)	238 (212–250)	>0.999
AST (U/L)	22 (18–24)	27 (25–36)	0.312
ALT (U/L)	20 (17–21)	22 (19–35)	0.343
BSQ (34–204)	83 (69.2–106)	66 (46.2–84.2)	0.177
BDI (0–63)	15 (12.5–19.8)	16 (11.2–17)	0.833
Pain	2 (33.3)	5 (83.3)	0.242
PGA EQ5D (0–100)	80 (53–100)	100 (100–100)	>0.999
IPAQ-SF			
Metabolic equivalent (MET/week)	2,050 (1,006–3,452)	2,876 (1,238–4,798)	0.844
Sedentary behavior (hours/day)	3.0 (2.1–5.8)	2.7 (2.0–4.6)	>0.999

Data are expressed as median (interquartile range) or frequency (%). ALT: alanine aminotransferase; AST: aspartate aminotransferase; ASyS: anti-synthetase syndrome; BDI: Beck Depression Inventory; BSQ: Body Shape Questionnaire; CPK: creatine phosphokinase; EQ5D: Quality of Life Questionnaire by EuroQol; FSS: Fatigue Severity Scale; HAQ: Health Assessment Questionnaire; IPAQ-SF: International Physical Activity Questionnaire—Short Form; LDH: lactate dehydrogenase; MET: metabolic equivalent; MMT: Manual Muscle Testing; MYOACT: Myositis Disease Activity Assessment Visual Analog Scale; PGA: Patient Global Assessment; PhGA: Physician Global assessment; SD: sexual dysfunction.

approaches in female SD[22-24]. Addressing sexual health through structured psychological interventions may therefore represent a valuable adjunct to standard rheumatologic care, particularly in conditions for which pharmacological options are limited.

Several limitations should be acknowledged. The single-center design and relatively small sample size may limit generalizability. In addition, the absence of a control group in the interventional phase precludes definitive conclusions regarding treatment efficacy. Nevertheless, this study represents the first investigation specifically focused on sexual health in ASyS and provides a foundation for future research. It is important to highlight that, given the high prevalence of SD in patients with ASyS and its direct impact on quality of life and, potentially, on adherence to treatment, routine screening for sexual health should be encouraged, as should the involvement of a multidisciplinary team (rheumatologists, pelvic floor physical therapists, psychologists, and sexual medicine specialists).

CONCLUSION

SD is highly prevalent in women with ASyS and appears to be intrinsically associated with ASyS, independent of disease activity or most psychosocial factors. The findings underscore the importance of systematically addressing sexual health in the clinical management of ASyS. Cognitive-behavioral therapy may improve sexual function, reduce sexual distress, and alleviate fatigue, representing a promising adjunctive strategy. Larger, multicenter, controlled studies are warranted to confirm these findings and guide evidence-based clinical interventions.

ETHICAL ASPECTS

This study was approved by the institutional Ethics Committee (opinion number 5.687.564 / CAAE 63366222.7.0000.0068) and registered in the Registro Brasileiro de Ensaios Clínicos—ReBEC (https://ensaiosclinicos.gov.br/) (number 14083). All participants signed written informed consent.

ACKNOWLEDGMENT

Maria Aurora Gomes da Silva for her value, support and dedication in assisting the group participating in cognitive behavioral therapy sessions. Marlise Sítima Mendes Simões Faria for her valuable partnership and ongoing encouragement during the data collection process.

AUTHORS' CONTRIBUTIONS

LRSS: Conceptualization, Data curation, Formal Analysis, Writing – original draft, Writing – review & editing. TL: Conceptualization, Data curation, Formal Analysis, Writing – original draft, Writing – review & editing. **EAGP:** Conceptualization, Data curation, Formal Analysis, Writing – original draft, Writing – review & editing. **AMS:** Conceptualization, Data curation, Formal Analysis, Writing – original draft, Writing – review & editing. **SKS:** 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. Katzap E, Barilla-LaBarca ML, Marder G. Anti-synthetase syndrome. Curr Rheumatol Rep. 2011;13:175-81. https://doi.org/10.1007/s11926-011-0176-8
2. Minopoulou I, Pyrgidis N, Tishukov M, Sokolakis I, Baniotopoulos P, Kefas A, et al. Sexual dysfunction in women with systemic autoimmune rheumatic disorders: a systematic review and meta-analysis. Rheumatology (Oxford). 2023;62:1021-30. https://doi.org/10.1093/rheumatology/keac457
3. Salonia A, Giraldi A, Chivers ML, Georgiadis JR, Levin R, Maravilla KR, et al. Physiology of women's sexual function: basic knowledge and new findings. J Sex Med. 2010;7:2637-60. https://doi.org/10.1111/j.1743-6109.2010.01810.x
4. American College of Obstetricians and Gynecologists' Committee on Practice Bulletins—Gynecology. Female sexual dysfunction: ACOG Practice Bulletin Clinical Management Guidelines for Obstetrician-Gynecologists, Number 213. Obstet Gynecol. 2019;134:e1-18. https://doi.org/10.1097/AOG.0000000000003324
5. Souza FHC, Araújo DB, Silva CA, Miossi R, Abdo CHN, Bonfá E, et al. Analysis of sexual function of patients with dermatomyositis and polymyositis through self-administered questionnaires: a cross-sectional study. Rev Bras Reumatol. 2017;57:134-40. https://doi.org/10.1016/j.rbre.2016.11.002
6. Greenfield J, Hudson M, Vinet E, Fortin PR, Bykerk V, Pineau CA, et al. A comparison of health-related quality of life (HRQoL) across four systemic autoimmune rheumatic diseases (SARDs). PLoS One. 2017;12:e0189840. https://doi.org/10.1371/journal.pone.0189840
7. Savel C, Cherillat MS, Berland P, Tronche AM, Soubrier M, Gerbaud L, et al. French survey on the crossed needs on sexual health for chronic inflammatory rheumatism patients and healthcare professionals. Rheumatol Int. 2020;40:1481-91. https://doi.org/10.1007/s00296-020-04630-4
8. Goldstein I, Kim NN, Clayton AH, DeRogatis LR, Giraldi A, Parish SJ, et al. Hypoactive sexual desire disorder: International

Society for the Study of Women's Sexual Health (ISSWSH) expert consensus panel review. Mayo Clin Proc. 2017;92(1):114-28. https://doi.org/10.1016/j.mayocp.2016.09.018

9. Heřmánková B, Špiritović M, Oreská S, Storkánová H, Komarc M, Klein M, et al. Sexual function in patients with idiopathic inflammatory myopathies: a cross-sectional study. Rheumatology (Oxford). 2021;60:5060-72. https://doi.org/10.1093/rheumatology/keab397
10. Behrens Pinto GL, Carboni RCS, Souza FHC, Shinjo SK. A prospective cross-sectional study of serum IL-17A in antisynthetase syndrome. Clin Rheumatol. 2020;39:2763-71. https://doi.org/10.1007/s10067-020-05013-7
11. Pacagnella RC, Vieira EM, Rodrigues OM, Souza C. Female sexual function index adaptation. Cad Saude Publica. 2008;24:416-26.
12. Pereira GMV, Juliato CRT, Gomes DAY, Beltramini TS, Monteiro MVC, Brito LGO, et al. Cross-cultural adaptation and validation of the Brazilian Portuguese version of the Female Sexual Distress Scale-Revised questionnaire for women with vaginal laxity. Int Urogynecol J. 2022;33:3163-70. https://doi.org/10.1007/s00192-022-05227-0
13. Abdo CHN. Female sexual quotient. RBM Rev Bras Med. 2006;63:477-82.
14. Pietro M, Silveira DX. Internal validity, dimensionality and performance of the Body Shape Questionnaire in a group of Brazilian college students. Rev Bras Psiquiatr. 2009;31:21-4. https://doi.org/10.1590/s1516-44462008005000017
15. Matsudo S, Araújo T, Matsudo V, Andrade D, Andrade E, Oliveira LC, et al. IPAQ validation. Rev Bras Ativ Fis Saude. 2001;6:5-18.
16. Bagattini AM, Camey SA, Miguel SR, Andrade MV, Noronha KVMS, Teixeira MAC, et al. Electronic version of the EQ-5D quality-of-life questionnaire: adaptation to a Brazilian population sample. Value Health Reg Issues. 2018;17:88-93. https://doi.org/10.1016/j.vhri.2017.11.002
17. Cruellas MG, Viana VS, Levy-Neto M, Souza FH, Shinjo SK. Myositis-specific and myositis-associated autoantibody profiles and their clinical associations in a large series of patients with polymyositis and dermatomyositis. Clinics. 2013;68:909-14. https://doi.org/10.6061/clinics/2013(07)04
18. Lerner T, Bagnoli VR, Pereyra EAG, Fonteles LP, Sorpreso ICE, Soares JM Jr, et al. Cognitive-behavioral group therapy for women with hypoactive sexual desire: a pilot randomized study. Clinics. 2022;77:100054. https://doi.org/10.1016/j.clinsp.2022.100054
19. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods. 2009;41:1149-60. https://doi.org/10.3758/BRM.41.4.1149
20. Rizzo ML. Statistical computing with R. 2nd ed. Boca Raton: Chapman & Hall/CRC; 2019.
21. Alidost F, Pakzad R, Dolatian M, Abdi F. Sexual dysfunction among women of reproductive age: a systematic review and meta-analysis. Int J Reprod Biomed. 2021;19:421-32. https://doi.org/10.18502/ijrm.v19i5.9251
22. Mestre-Bach G, Blycker GR, Potenza MN. Behavioral therapies for treating female sexual dysfunctions: a state-of-the-art review. J Clin Med. 2022;11:2794. https://doi.org/10.3390/jcm11102794
23. Brotto LA, Zdaniuk B, Chivers ML, Jabs F, Grabovac A, Lalumière ML, et al. A randomized trial comparing group mindfulness-based cognitive therapy with group supportive sex education and therapy for the treatment of female sexual interest/arousal disorder. J Consult Clin Psychol. 2021;89:626-39. https://doi.org/10.1037/ccp0000661
24. Sharifipour F, Qaderi K, Peighambardoost R, Zahedian M, Azizi F, Siahkal SF. Impact of cognitive behavio

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260531

Intrapartum rectovaginal distance as a predictor of obstetric anal sphincter injury: a prospective cohort study

Elif Görkem Bademci[1]*, Yağmur Şimşek Erdim[2], Ecem Yüksel Özsoy[3], Erkan Elçi[4]

SUMMARY

OBJECTIVE: The aim of this study was to evaluate the relationship between intrapartum rectovaginal distance measured by transperineal ultrasonography and obstetric anal sphincter injury severity and second-stage labor duration in nulliparous women undergoing vaginal delivery.

METHODS: This prospective cohort study was conducted at a tertiary referral center in Istanbul, between September 2021 and January 2022. Of 200 consecutive nulliparous women with singleton pregnancies at ≥37 weeks' gestation, 148 who underwent vaginal delivery were included in the primary analysis. Intrapartum rectovaginal distance was measured in the midsagittal plane by 2D transperineal ultrasonography using a single operator during the inter-contraction rest period, at a fetal head station ranging from -1 to +1. Angle of progression and levator ani anteroposterior diameter were also recorded. Obstetric injury was graded per American College of Obstetricians and Gynecologists 2016 criteria; grade 3–4 laceration (severe obstetric anal sphincter injury) was the primary outcome.

RESULTS: Severe obstetric anal sphincter injury occurred in 18 vaginal deliveries (12.2%). Rectovaginal distance was significantly shorter in women with severe obstetric anal sphincter injury (6.91±1.33 mm vs. 9.94±2.75 mm, p<0.001). Second-stage duration was longer with severe obstetric anal sphincter injury (125.8±6.5 vs. 91.3±18.0 min, p<0.001). Receiver operating characteristic analysis identified an rectovaginal distance cut-off of 8.1 mm (AUC=0.846; sensitivity 88.9%, specificity 80.8%). On multivariate logistic regression, rectovaginal distance (OR 0.516, 95%CI 0.373–0.713) and body mass index (OR 1.280, 95%CI 1.078–1.521) were independently associated with severe obstetric anal sphincter injury.

CONCLUSION: Intrapartum rectovaginal distance measurement by transperineal ultrasonography is a noninvasive, readily obtainable parameter that independently predicts severe obstetric anal sphincter injury and may support individualized intrapartum risk stratification. Multicenter validation with larger samples is warranted.

KEYWORDS: Obstetric labor complications. Perineum. Ultrasonography. Parity. Labor stage. Second.

INTRODUCTION

Obstetric anal sphincter injury (OASI) represents the most clinically consequential complication of vaginal delivery, with an incidence of approximately 3% across all births and nearly 6% among nulliparous women[1]. Affected women face substantially elevated risks of fecal incontinence, dyspareunia, and chronic pelvic pain, with symptoms frequently remaining latent for decades[2-4]. A short perineal body, instrumental delivery, and nulliparity have been identified as significant anatomical and obstetric risk factors[5]. Despite growing awareness, prospective identification of at-risk women before delivery has yet to be incorporated into routine practice.

Transperineal ultrasonography (TPUS) offers non-invasive, real-time evaluation of fetal head descent and pelvic floor parameters[6]. The rectovaginal distance (RVD)—the shortest sonographic interval between the posterior vaginal wall and the rectal mucosa—has emerged as a surrogate for perineal tissue reserve. Prior reports linking reduced RVD to higher-grade trauma derive predominantly from postpartum examinations, where edema, hematoma, and surgical repair confound tissue geometry[7,8]. Intrapartum measurement during the active second stage may therefore yield superior predictive validity.

This study aimed to evaluate the relationship between intrapartum RVD, measured by 2D TPUS, and severe OASI in nulliparous women undergoing vaginal delivery, and to establish a sonographic cut-off for clinical risk stratification. Secondary objectives included assessment of the associations between angle of progression (AoP), levator ani contractile reserve, and second-stage labor duration.

[1]Sancaktepe Sehit Prof. Dr. Ilhan Varank Training and Research Hospital, Department of Gynecology and Obstetrics – Istanbul, Turkey.
[2]Sisli Hamidiye Etfal Training and Research Hospital, Department of Gynecology and Obstetrics – Istanbul, Turkey.
[3]Darıca Farabi Training and Research Hospital, Department of Gynecology and Obstetrics – Kocaeli, Turkey.
[4]Medicana Atakoy Hospital, Department of Gynecology and Obstetrics – Istanbul, Turkey.
*Corresponding author: elifbayazitli@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on May 29, 2026. Accepted on August 15, 2026.
Scientific Editor: José Maria Soares Júnior

METHODS

This prospective cohort study enrolled consecutive nulliparous women (singleton, ≥37 weeks) at a tertiary referral hospital between September 2021 and January 2022. Ethical approval was obtained (Istanbul Ümraniye Training and Research Hospital Ethics Committee, Decision No: 15878), and written informed consent was provided by all participants.

Inclusion criteria: nulliparous women aged 18–45 years with confirmed gestational age. Exclusion criteria: chronic disease; BMI >35 kg/m^2; multiple pregnancy; prior uterine surgery; substance use; fetal anomalies; assisted reproductive technology (ART) conception; gestational age <37 weeks; non-cephalic presentation. Eligible women were consecutively enrolled and data recorded prospectively.

Ultrasound evaluation

All TPUS measurements were performed by a single experienced operator [E.E.] using an Esaote MyLab Seven ultrasound system (Esaote S.p.A., Genova, Italy) equipped with a convex array transducer (SB2C41, 1–8 MHz), with the patient in the lithotomy position. RVD was measured in the midsagittal plane as the shortest distance between the rectal mucosa and the posterior vaginal wall at the anorectal junction during the second stage; the mean of three measurements was used. All measurements were obtained during the inter-contraction rest period at a fetal head station ranging from -1 to +1 (mean: -0.4±0.7), ensuring reproducibility and reflecting resting tissue thickness. For the purposes of receiver operating characteristic (ROC) analysis, the optimal RVD threshold for prediction of severe OASI was determined post hoc using the Youden index. AoP was defined as the angle between the long axis of the symphysis pubis and a tangential line to the deepest bony part of the fetal head, measured in the midsagittal plane[6]. An AoP of ≥120° was considered indicative of imminent delivery, while values <120° were classified as unfavorable for progression, based on previously published thresholds. Levator ani muscle function was assessed by measuring the anteroposterior diameter of the levator hiatus at rest and during maximal voluntary contraction in the midsagittal plane. Contraction strength was expressed as the proportional reduction in hiatal diameter: (rest diameter - contracted diameter)/rest diameter×100 (%)[9]. A reduction of ≥20% was considered indicative of adequate contractile function. Labor duration was defined as the interval from full cervical dilation (10 cm) to delivery of the fetus, corresponding to the second stage of labor.

Active pushing with Valsalva maneuvers was used throughout the second stage under continuous fetal heart rate monitoring. Hands-on perineal support was applied and episiotomy performed at the clinician's discretion. Perineal trauma was assessed clinically immediately following delivery and confirmed by postpartum TPUS, both performed by the same operator who conducted the intrapartum measurements. Injuries were classified per American College of Obstetricians and Gynecologists (ACOG) 2016 criteria. Severe OASI was defined as third- or fourth-degree injury and constituted the primary outcome. Episiotomy sites were assessed separately and excluded from the laceration grading system.

Sample size and statistical analysis

The sample comprised consecutively available patients; post hoc power analysis (G*Power 3.1) estimated 94% power (α=0.05, Cohen's d=0.8). Continuous variables are reported as mean±standard deviation (SD) or median [interquartile range (IQR)]; categorical variables as n (%). Group comparisons used the Student t-test or Mann-Whitney U test, and chi-square or Fisher's exact test as appropriate. Multivariate logistic regression with severe OASI as the outcome was performed; variables with p<0.05 in univariate analysis were considered for inclusion, with a maximum of two predictors per the events-per-variable rule (18 events). Discriminative ability was assessed by ROC analysis with a Youden index–derived threshold. Significance: p<0.05.

RESULTS

Of 272 women assessed, 72 were excluded (36 chronic disease, 24 BMI >35 kg/m^2, 12 multiple pregnancy) and 200 enrolled. Of these, 52 cesarean deliveries were excluded; the final cohort comprised 148 vaginal deliveries. Mean maternal age was 25.0±5.6 years, gestational age 39.0±1.4 weeks, and BMI 28.0±3.7 kg/m^2. Episiotomy was performed in 133 women (89.9%), and operative vaginal delivery (Kiwi Omnicup) in 13 (8.8%). Mean intrapartum RVD was 9.6±2.8 mm, AoP 108.1±13.4°, and resting levator ani diameter 52.1±8.9 mm.

Perineal trauma occurred in 130 women (87.8%), including 18 (12.2%) with severe OASI (11 third-degree, 7 fourth-degree). Severe OASI (grade 3–4) occurred in 18 women (12.2%). BMI, RVD, AoP, resting levator ani diameter, second-stage duration, Apgar scores, and operative delivery rate differed significantly between women with and without severe OASI (all p<0.05; Table 1). Gestational age, BPD, birth weight, levator ani contraction strength, and episiotomy rate were comparable between groups.

RVD demonstrated high discriminative ability for predicting severe OASI (AUC=0.846); the optimal cut-off was 8.1 mm (sensitivity 88.9%, specificity 80.8%) (Figure 1). On multivariate logistic regression, RVD (OR 0.516, 95%CI

0.373–0.713, p<0.001) and BMI (OR 1.280, 95%CI 1.078–1.521, p=0.005) were identified as independent predictors of severe OASI (Table 2).

DISCUSSION

This study identified a clinically applicable intrapartum RVD threshold for severe OASI risk stratification. ROC analysis yielded an AUC of 0.846, with a cut-off of 8.1 mm achieving 88.9% sensitivity and 80.8% specificity. These metrics position intrapartum RVD as a useful screening parameter. The severe OASI rate of 12.2% observed in this cohort exceeds the 3–6% reported in general obstetric populations; this is attributable to the exclusively nulliparous study design, the high rate of operative vaginal delivery (66.7% in the severe OASI group), and the use of TPUS for injury detection, which has been shown to identify sphincter defects missed by clinical examination alone[10,11].

Prior studies have linked reduced anovaginal tissue thickness to higher-grade perineal trauma; Pihl et al. and Hjertberg et al. demonstrated inverse relationships between RVD and OASI severity[7,8]. A 2023 meta-analysis by Packet et al. further identified perineal tissue dimensions and delivery mode as predictive factors[5]. The present study extends these observations by measuring RVD during active labor, avoiding the measurement variability introduced by postpartum edema and hematoma[10,12], and establishes a precise numerical threshold for clinical use.

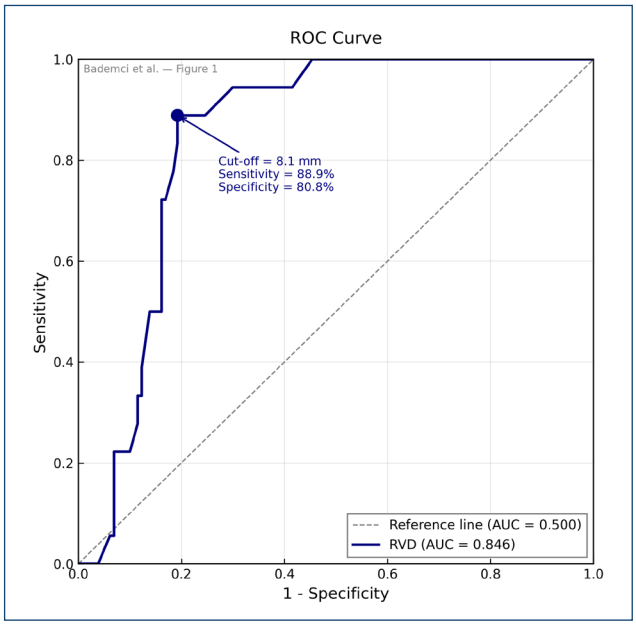


Figure 1. Receiver operating characteristic curve for intrapartum rectovaginal distance in predicting severe obstetric anal sphincter injury (grade 3–4). The optimal cut-off of 8.1 mm was determined by the Youden index (AUC=0.846; sensitivity 88.9%, specificity 80.8%). AUC: area under the curve; RVD: rectovaginal distance.

Table 1. Comparison of clinical, ultrasonographic, and neonatal characteristics between women with severe obstetric anal sphincter injury (grade 3–4) and those without severe injury (grade 0–2) among 148 nulliparous women undergoing vaginal delivery.

Variable	Severe OASI (n=18)	No/mild OASI (n=130)	p-value
Maternal age (years)	26.1±5.6	24.8±5.6	0.313
Gestational age (weeks)	38.6±1.0	39.0±1.4	0.117
BMI (kg/m²)	29.9±2.5	27.8±3.7	0.005*
Rectovaginal distance (mm)	6.91±1.33	9.94±2.75	<0.001*
Angle of progression (°)	102.2±10.2	108.9±13.7	0.006*
Rest levator ani (mm)	55.7±5.2	51.6±9.2	0.011*
Levator ani contraction (mm)	42.1±2.7	43.7±10.7	0.522
BPD (mm)	91.6±3.2	91.4±3.4	0.962
Birth weight (g)	3379±413	3196±402	0.053
Second-stage duration (minutes)	125.8±6.5	91.3±18.0	<0.001*
1st min APGAR	5.6±1.0	7.0±0.8	<0.001*
5th min APGAR	7.1±0.8	8.3±0.8	<0.001*
Episiotomy, n (%)	18 (100%)	115 (88.5%)	0.217
Operative delivery (Kiwi), n (%)	12 (66.7%)	1 (0.8%)	<0.001*

Data are presented as mean±standard deviation or number (percentage). Between-group comparisons were performed using the Mann-Whitney U test for continuous variables and Fisher's exact test for categorical variables. Statistically significant results are indicated with an asterisk (*p<0.05). Second-stage duration was defined as the interval from full cervical dilation to delivery of the fetus. BMI: body mass index; RVD: rectovaginal distance; AoP: angle of progression; BPD: biparietal diameter; APGAR: Apgar score; OASI: obstetric anal sphincter injury.

Table 2. Univariate and multivariate logistic regression analysis for predictors of severe obstetric anal sphincter injury (grade 3–4) among 148 nulliparous women undergoing vaginal delivery.

Variable	OR	95%CI	p-value
Panel A—Univariate analysis			
RVD (mm)	0.594	0.461–0.765	<0.001*
BMI (kg/m²)	1.165	1.020–1.330	0.025*
Second-stage duration (minutes)	1.542	1.194–1.991	0.001*
Operative delivery	258.0	28.6–2324.0	<0.001*
AoP (°)	0.962	0.925–1.000	0.051
Rest levator ani (mm)	1.055	0.996–1.117	0.069
Birth weight (g)	1.001	1.000–1.003	0.075
Levator ani contraction (mm)	0.984	0.937–1.034	0.528
Panel B—Multivariate logistic regression (Nagelkerke R^2=0.362)			
RVD (mm)	0.516	0.373–0.713	<0.001*
BMI (kg/m²)	1.280	1.078–1.521	0.005*

Panel A presents crude odds ratios from univariate analysis; Panel B presents adjusted odds ratios from multivariate logistic regression including variables with p<0.05 on univariate analysis, limited to two predictors per the events-per-variable rule (18 events). Results are expressed as odds ratios (ORs) with 95%CIs. Statistically significant results are indicated with an asterisk (*p<0.05). Nagelkerke R^2=0.362. BMI: body mass index; RVD: rectovaginal distance; AoP: angle of progression; OR: odds ratio; OASI: obstetric anal sphincter injury; CI: confidence interval.

For women with an intrapartum RVD below 8.1 mm, evidence-based strategies including controlled fetal head delivery, hands-on perineal support, and selective mediolateral episiotomy may reduce sphincter injury risk[11,13]. Conversely, women with RVD exceeding 8.1 mm may not require routine episiotomy, consistent with current guidelines[11,14]. A 2025 systematic review confirmed that 2D TPUS in the delivery room has utility in assessing labor progress and recognizing OASI[15].

A secondary aim of this study was to characterize the relationship between AoP and labor dynamics. In contrast to prior reports linking wider AoP to shorter labor duration[16,17], AoP was not significantly associated with second-stage duration in our cohort (Spearman r=-0.119, p=0.149). This discrepancy may reflect the narrow range of AoP values observed during the active second stage, as measurements were obtained at a fetal head station ranging from -1 to +1, limiting variability in descent angle. However, AoP was significantly lower in women with severe OASI compared to those without (102.2° vs. 108.9°, p=0.006), suggesting that a more acute angle of fetal descent may be associated with greater perineal tissue loading, though this association did not reach independence in multivariate analysis. Descent mechanics and perineal tissue vulnerability thus appear to represent partially overlapping but distinct physiological domains.

Levator ani contractile function did not differ between women with and without severe OASI (p=0.522), consistent with Volløyhaug et al., who similarly found no early link between levator ani morphology and continence outcomes postpartum[18]. Pelvic floor dysfunction following OASI is recognized to follow a delayed trajectory, often manifesting decades after delivery as neurological and hormonal reserves decline[3].

The limitations of the study include its single-center design, limited sample size, short follow-up period, and use of 2D measurements. Clinical anovaginal distance was not systematically recorded, precluding a direct comparison with sonographic RVD. The high institutional episiotomy rate (89.9%) reflects local practice patterns and limits generalizability to settings with selective episiotomy policies. Furthermore, the sample size, while adequately powered for large effect sizes, precludes construction of a robust multivariable prediction model. Notwithstanding these limitations, the prospective design, reduced group heterogeneity, and derivation of an RVD cut-off constitute the study's major strengths. Additionally, both intrapartum RVD measurement and postpartum OASI assessment were performed by the same operator, introducing the possibility of observer bias; blinded outcome assessment would strengthen future studies.

CONCLUSION

Intrapartum RVD measured by TPUS is a non-invasive, readily obtainable parameter that independently predicts severe OASI and may support individualized risk stratification during the second stage of labor. Routine intrapartum RVD assessment may complement existing clinical tools for perineal injury prevention in nulliparous women. Multicenter validation with larger samples is warranted before widespread clinical implementation.

AUTHORS' CONTRIBUTIONS

EGB: Methodology, Writing – original draft, Writing – review & editing. **YŞE:** Conceptualization, Formal Analysis. E**YÖ:** Data curation. **EE:** 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. Orlando A, Thomas G, Murphy J, Hotouras A, Bassett P, Vaizey C. A systematic review and a meta-analysis on the incidence of obstetric anal sphincter injuries during vaginal delivery. Colorectal Dis. 2024;26:227-42.
2. Nilsson IEK, Åkervall S, Molin M, Milsom I, Gyhagen M. Severity and impact of accidental bowel leakage two decades after no, one, or two sphincter injuries. Am J Obstet Gynecol. 2023;228:447.e1-19.
3. Walsh M, Fitzpatrick M, Anglim B. Obstetric anal sphincter injury—the long game: primary and cumulative anal sphincter injury during childbirth and its long-term implications—a clinical practice review. Gynecol Pelvic Med. 2023;6:5.
4. DeLancey JOL, Masteling M, Pipitone F, LaCross J, Mastrovito S, Ashton-Miller JA. Pelvic floor injury during vaginal birth is life-altering and preventable: what can we do about it? Am J Obstet Gynecol. 2024;230:279-94.
5. Packet B, Page AS, Cattani L, Bosteels J, Deprest J, Richter J. Predictive factors for obstetric anal sphincter injury in primiparous women: systematic review and meta-analysis. Ultrasound Obstet Gynecol. 2023;62:486-96.
6. Kalache KD, Dückelmann AM, Michaelis SA, Lange J, Cichon G, Dudenhausen JW. Transperineal ultrasound imaging in prolonged second stage of labor with occipitoanterior presenting fetuses: how well does the 'angle of progression' predict the mode of delivery? Ultrasound Obstet Gynecol. 2009;33:326-30.
7. Pihl S, Uustal E, Blomberg M. Anovaginal distance and obstetric anal sphincter injury: a prospective observational study. Int Urogynecol J. 2019;30:939-44.
8. Hjertberg L, Uustal E, Pihl S, Blomberg M. Maternal body mass ındex and anovaginal distance in active phase of term labor. Biomed Res Int. 2018;2018:1532949.
9. Nyhus MØ, Oversand SH, Salvesen Ø, Salvesen KÅ, Mathew S, Volløyhaug I. Ultrasound assessment of pelvic floor muscle contraction: reliability and development of an ultrasound-based contraction scale. Ultrasound Obstet Gynecol. 2020;55:125-31.
10. Stickelmann AL, Kennes LN, Hölscher M, Graef C, Kupec T, Wittenborn J, et al. Obstetric anal sphincter injuries (OASIS): using transperineal ultrasound (TPUS) for detecting, visualizing and monitoring the healing process. BMC Womens Health. 2022;22:339.
11. Okeahialam NA, Wong KW, Jha S, Sultan AH, Thakar R. Mediolateral/lateral episiotomy with operative vaginal delivery and the risk reduction of obstetric anal sphincter injury (OASI): a systematic review and meta-analysis. Int Urogynecol J. 2022;33:1393-405.
12. Bellussi F, Dietz HP. Postpartum ultrasound for the diagnosis of obstetrical anal sphincter injury. Am J Obstet Gynecol MFM. 2021;3:100421.
13. Okeahialam NA, Sultan AH, Thakar R. The prevention of perineal trauma during vaginal birth. Am J Obstet Gynecol. 2024;230:S991-1004.
14. Committee on Practice Bulletins-Obstetrics. ACOG Practice Bulletin No. 198: prevention and management of obstetric lacerations at vaginal delivery. Obstet Gynecol. 2018;132:e87-102.
15. Nebel S, Vardon D, Dreyfus M, Pizzoferrato AC. 2D-transperineal ultrasound in delivery room: contribution in assessing labor progress, predicting outcome of labor and recognizing obstetric anal sphincter injuries (OASIS). A systematic review. J Gynecol Obstet Hum Reprod. 2025;54:102910.
16. Youssef A, Brunelli E, Montaguti E, Donna G, Dodaro MG, Bianchini L, et al. Transperineal ultrasound assessment of maternal pelvic floor at term and fetal head engagement. Ultrasound Obstet Gynecol. 2020;56:921-7.
17. Jung JE, Lee YJ. Intrapartum transperineal ultrasound: angle of progression to evaluate and predict the mode of delivery and labor progression. Obstet Gynecol Sci. 2024;67:1-16.
18. Volløyhaug I, Taithongchai A, Gruting I, Sultan A, Thakar R. Levator ani muscle morphology and function in women with obstetric anal sphincter injury. Ultrasound Obstet Gynecol. 2019;53:410-6.

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260580

Moral distress arising from professionalism dilemmas among medical interns

Giulia Francci[1], Laura Oliveira de Alcântara Ferraz[1], Marina da Rocha Brito[1], Rogério Aparecido Dedivitis[1]*

SUMMARY

BACKGROUND: Moral distress is defined as a psycho-emotional experience that occurs when individuals recognize the morally appropriate course of action but are unable to carry it out due to external constraints such as institutional hierarchies, organizational norms, or authority pressures. In medical education, this phenomenon is particularly frequent, given the deeply ethical nature of clinical decisions and patient care.

OBJECTIVE: The aim of this study was to translate and adapt the Moral Distress from Professionalism Dilemmas checklist into Brazilian Portuguese.

METHODS: The translation and cross-cultural adaptation process followed internationally recognized guidelines. It included independent translations by two professionals (one with and one without medical training), synthesis of the versions, back-translation, expert committee review, pretesting with 200 medical interns, and final evaluation.

RESULTS: Of the 200 medical students invited, 160 (80%) completed the checklist, with no significant differences by year or gender. Depending on the item, between 0 and 57.5% of participants reported exposure to the professionalism dilemmas investigated. Among exposed students, the intensity of moral distress varied considerably across situations.

CONCLUSION: Moral distress is a prevalent phenomenon in medical education. The adapted checklist demonstrated semantic and conceptual equivalence to the original and may serve as a valuable tool to identify, quantify, and understand moral distress among medical students in clinical training.

KEYWORDS: Moral distress. Medical professionalism. Ethical dilemmas. Medical education.

INTRODUCTION

Moral distress is a psycho-emotional experience that arises when an individual recognizes the ethically appropriate action but feels constrained from acting upon it due to external factors such as institutional hierarchies, organizational rules, or pressures from authority figures[1]. In medicine, this phenomenon is frequent because clinical decision-making and patient care are inherently moral activities. The recurrent dissonance between what one believes to be right and what can be done in practice may lead to frustration, guilt, loss of empathy, and, in the long term, burnout[2,3].

Within this context, what is taught about professional conduct and ethics throughout medical education often becomes a source of tension. Physicians have an ethical commitment to integrity, responsibility, empathy, technical competence, and respect toward patients—principles that should be consistently applied and modeled during clinical training. Medical interns, typically in the final years of undergraduate education, occupy the lowest level in the hospital hierarchy and are therefore in a position of considerable moral vulnerability. They frequently witness unprofessional behaviors by peers or superiors but feel powerless to intervene, question, or even express concern. Such circumstances foster ethical dilemmas and contribute to moral distress, which in turn affects not only interns' mental health but also their professional and moral development[4].

Understanding moral distress resulting from professionalism dilemmas experienced by medical interns is therefore essential for improving pedagogical strategies in medical education and for promoting healthier, more ethical, and more formative clinical learning environments.

The Moral Distress from Professionalism Dilemmas checklist[5] investigates the frequency of events in which ethical values conflict with institutional norms. It addresses topics such as hierarchical relationships, inappropriate behaviors, workplace pressures, privacy concerns, and organizational limitations. By quantifying both the occurrence and the perceived impact of such experiences, the instrument allows for analysis of moral distress patterns and their correlations with social and academic

[1]Centro Universitário Lusíada – Santos (SP), Brazil.
*Corresponding author: dedivitis@usp.br
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on June 03, 2026. Accepted on August 08, 2026.
Scientific Editor: Roseli Nomura

variables. Moreover, it supports the development of pedagogical interventions aimed at fostering ethical, safe, and emotionally sustainable clinical environments.

There is a lack of an instrument adapted specifically for medical interns and, as a result, it is difficult measuring and monitoring educational interventions. Thus, this study aimed to translate and adapt the Moral Distress from Professionalism Dilemmas checklist into Brazilian Portuguese among fifth- and sixth-year medical students, to assess moral distress related to professionalism dilemmas during internship training.

METHODS

This study was previously approved by the institutional review board of the institution where it was performed and all participants who accept the participation signed digitally the free and informed consent form.

Following consent and authorization from the original instrument's author, the translation and cross-cultural adaptation process was conducted in accordance with established international guidelines[6]. The procedure comprised six sequential stages: (1) initial translation, (2) synthesis of translations, (3) back-translation, (4) expert committee review, (5) pretesting of the pre-final version, and (6) final assessment of the entire adaptation process.

Stage I: Inicial translation: The first stage involved two independent forward translations from English (source language) into Brazilian Portuguese (target language). Each translator had a distinct professional background, allowing for comparison of perspectives, identification of ambiguities, and detection of inadequate wording for further discussion and resolution. Both translators submitted written reports justifying their lexical and syntactic choices and noting any uncertainties or challenging phrases. Translator 1: was a medical professional familiar with the concepts addressed in the instrument, providing a clinically oriented translation. Translator 2: was a professional without a medical background and was not informed about the checklist's purpose, ensuring a linguistically natural translation unaffected by disciplinary bias.

Stage II: Synthesis of translations: A synthesis meeting was held among both translators and an observer to reconcile discrepancies and produce a single synthesized version (T-12). A report documenting all decisions, discussions, and rationales for the final choices was prepared at the end of this stage.

Stage III: Back-translation: Two translators, native English speakers blind to the original instrument, independently back-translated the synthesized Brazilian Portuguese version (T-12) into English to verify the accuracy and conceptual fidelity of the translation. These translators had no medical background or prior knowledge of the checklist's content. Two back-translated versions (BT1 and BT2) were produced and compared with the original instrument to detect any conceptual or semantic inconsistencies.

Stage IV: Expert committee review: An expert committee—composed of two health professionals, three undergraduate medical students, two native English translators, and back-translators involved in the process—reviewed all prior documents to reach consensus on discrepancies and ensure conceptual equivalence. The committee's evaluation focused on four main areas: semantic, idiomatic, experiential, and conceptual correspondence. This review resulted in the pre-final version of the instrument in Brazilian Portuguese. During the expert committee review, particular attention was given to preserving conceptual equivalence while ensuring contextual relevance. Whenever an item was considered potentially inapplicable to Brazilian undergraduate medical training, the committee evaluated whether adaptation, replacement, or exclusion would best preserve the underlying construct. Decisions were based on consensus and aimed to minimize threats to the content validity of the adapted instrument while maintaining maximum comparability with the original version.

Stage V: Pretesting of the pre-final version: The pre-final version of the Moral Distress from Professionalism Dilemmas checklist was administered to fifth- and sixth-year medical interns from the institution. Cognitive interviews or debriefing (e.g., think-aloud procedures, comprehension and acceptability probing) were not conducted as pretest, since the content of the checklist was considered intuitive. The instrument was distributed electronically via Google Forms following completion of the informed consent form. A total of 200 students was invited to participate voluntarily. Each participant completed the checklist individually and anonymously in a single sitting. The checklist evaluates two related but distinct constructs. First, participants indicate whether they have been exposed to each professionalism dilemma. Second, for situations that occurred, they rate the intensity of the resulting moral distress using a four-point scale. Thus, frequency of exposure reflects the occurrence of professionalism dilemmas, whereas the distress score reflects participants' emotional response to those events rather than their frequency. Responses were scored on a four-point moral distress scale: 1—no distress (although it happened); 2—mild distress (slightly uncomfortable during the event, but not much thought afterward); 3—moderate distress (disturbed and distressed during the event, felt helpless to do anything and thought about it for a while afterward, but does

not bother me much now); and 4—severe distress (disturbed and distressed during the event, felt helpless to do anything, and still bothers me). There were not domains or subscales in the checklist.

Stage VI: Final evaluation: A comprehensive review of the entire translation and adaptation process was conducted by the authors to confirm linguistic accuracy and conceptual equivalence across all stages.

Statistical analysis: Categorical variables were described using frequency distributions (number and percentage), while numerical data were analyzed using measures of central tendency (mean and median) and variability (range and standard deviation). To compare total checklist scores between fifth- and sixth-year medical students, the Mann-Whitney U test was applied. Associations between total scores, academic year, and gender were assessed using the Kruskal-Wallis test. A boxplot was generated to visually represent associations between numerical and categorical variables. A 5% significance level was adopted for all statistical tests. All analyses were performed using Stata Statistical Software, version 19[7]. In the comparison between the groups of students, the total score with the average of item responses were considered.

RESULTS

During the process of translation and cultural adaptation of the checklist, several minor adjustments were made to ensure greater clarity and appropriateness to the Brazilian academic and cultural context.

The main adaptation involved substituting the term "resident" with "intern" (items S3 and S53), as the target population of this study comprised undergraduate medical interns rather than postgraduate residents. Similarly, "medical residency" was translated as "medical internship" (item S4) to maintain contextual coherence.

The expressions "senior physicians" were adapted to "supervisors" (items S11 and S20) or "medical supervisors" (item S29) to better reflect the hierarchical and educational structures of Brazilian medical training.

Additional lexical refinements were made to improve readability and conceptual precision. For instance, "alcohol" was rendered as "alcoholic beverages" (item S6), and "being caught" was adapted to "being found" (item S7). The term "nepotism" was replaced by "favoritism in scheduling or providing medical services to relatives or friends" (item S44) to prevent potential ambiguity. Similarly, "red envelope"—a culturally specific expression—was adapted to "gratuities" (item S47) to achieve Brazilian cultural equivalence.

Item S18, which addresses "salary and benefits," was excluded because this situation is not applicable to undergraduate medical interns in Brazil, who do not receive salaries or employment-related benefits during their clinical training. Rather than reflecting a cultural or linguistic difference, this exclusion results from a structural characteristic of Brazilian medical education, in which internship constitutes a compulsory curricular component rather than an employment relationship. The expert committee considered retaining the item with modified wording; however, any adaptation would substantially alter its original construct, potentially compromising content validity. Therefore, exclusion was judged to be the most appropriate strategy to preserve conceptual integrity of the remaining instrument. Consequently, the Brazilian version contains one fewer item than the original checklist, and direct score comparisons with studies using the original instrument should take this structural difference into account. Finally, items S23 and S48 underwent minor linguistic adjustments for clarity, without altering their original meaning (Supplementary 1).

Among the 200 students invited to participate, 160 (80%) completed the checklist. Depending on the checklist item, between 0 and 57.5% of participants reported having experienced the situations described. Among those exposed to each situation, the intensity of moral distress varied according to the four-point response scale, with substantial heterogeneity across items (Figure 1).

When comparing results by academic year (fifth vs. sixth year) and gender, no statistically significant differences were observed (Table 1).

DISCUSSION

Analysis of the 57 situations included in the Moral Distress from Professionalism Dilemmas checklist revealed that moral distress is present among medical interns, although with varying intensities. The predominant categories reported were mild stress (38%) and moderate stress (34%), while no stress accounted for approximately 18% and severe stress for around 10% of responses. This pattern suggests that, even though not all situations provoke intense suffering, a consistent background level of moral discomfort permeates students' clinical experiences.

It is important to distinguish between exposure to professionalism dilemmas and the emotional impact resulting from those experiences. Frequent exposure does not necessarily produce severe moral distress, and conversely, relatively uncommon events may generate intense moral suffering when they involve profound ethical conflicts. Therefore, the checklist should be interpreted as assessing both the occurrence of

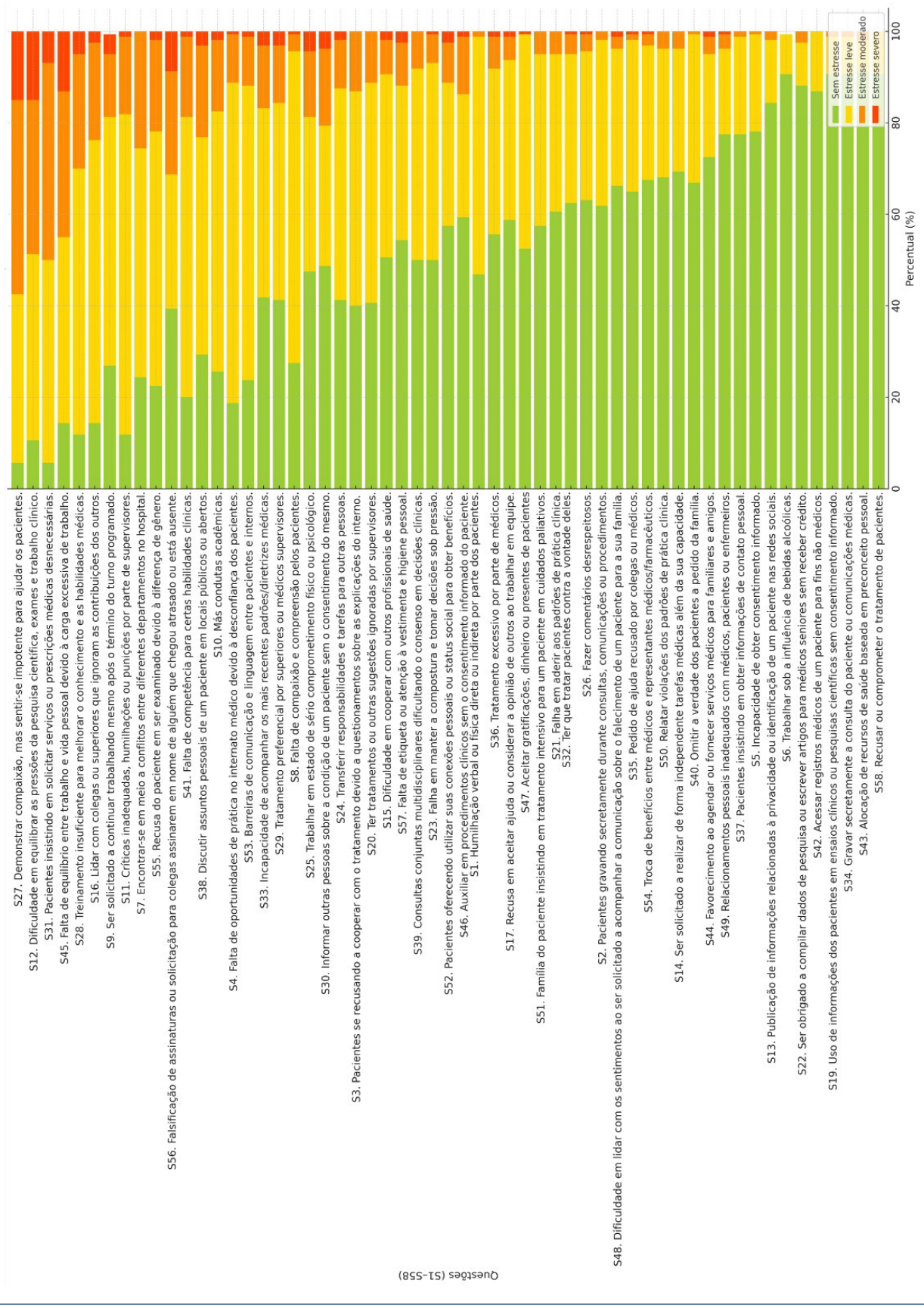


Figure 1. Intensity of moral distress related to professionalism dilemmas.

Table 1. Total questionnaire score according to academic year (fifth and sixth year) and gender.

Score measures	Total	Fifth year		Sixth year		p-value
		Male	Female	Male	Female	
		Measures				
n	160	31	58	24	47	0.847
Range	92–249	142–249	148–249	159–229	92–220	
Median	186	184.0	182.5	188.5	187.0	
Mean	185.8	184.8	185.6	188.2	185.7	
Standard deviation	22.2	23.3	21.7	18.3	24.4	

p-value obtained using the Kruskal-Wallis test.

ethically challenging situations and the magnitude of the distress they provoke, rather than assuming a direct relationship between frequency and emotional impact.

The exclusion of item S18 deserves consideration when interpreting cross-cultural equivalence. Although the adaptation preserved the conceptual domains explored by the original checklist, one employment-related situation could not be meaningfully represented in the Brazilian undergraduate context. Consequently, conceptual equivalence was maintained at the instrument level, whereas strict structural equivalence was slightly modified by reducing the total number of items. This difference is unlikely to affect the assessment of professionalism-related moral distress among Brazilian medical interns because the excluded scenario is absent from their educational reality. Nevertheless, future international comparative studies should acknowledge this difference when comparing total scores or psychometric properties across versions, preferably using standardized or item-level analyses rather than raw total scores.

Clinical care involves the interaction of multiple moral agents—including patients, families, and healthcare professionals—each of whom holds personal perspectives on what is right or wrong, desirable or undesirable, and assigns relative weight to these moral values[1]. The moral distress experienced by healthcare professionals typically escalates during crises and diminishes once the precipitating event is resolved. However, a moral residue remains, contributing to an elevated baseline level of distress that amplifies during subsequent crises[8].

The concept of moral distress is relatively recent, with approximately three decades of scholarly discussion. Most works on the topic cite Jameton, who introduced the term to distinguish moral distress from other types of ethical problems in healthcare. Situations that generate moral distress are those in which the professional clearly understands which moral principle applies but is constrained from acting accordingly. When uncertainty exists regarding which principle should guide action, the situation constitutes moral uncertainty rather than moral distress. Furthermore, moral distress differs from a moral dilemma: in the former, the individual knows the correct action but is unable to perform it; in the latter, two or more moral principles conflict, each supporting different actions[9].

The situations that elicited the highest levels of moderate to severe stress in this study were those related to medical hierarchy, ethical limitations in clinical practice, and conflicts of values—such as verbal humiliation, lack of autonomy in decision-making, and pressure to follow directives contrary to one's moral convictions. These findings are consistent with those of Song et al., who also identified hierarchical and authority-related dilemmas as leading sources of moral distress among medical trainees.

The comparative analysis revealed no significant associations between moral distress levels and gender or academic year, suggesting that distress is shaped more by shared institutional contexts than by individual demographic factors. Similar findings have been reported among medical residents, who, despite being at different training stages, share similar perceptions regarding professional and unprofessional conduct, indicating that hospital environments tend to homogenize moral perceptions[4].

Conversely, women approach moral issues through an ethic of care, whereas men are guided more by principles of justice[10]. This distinction could theoretically influence perceptions of moral distress. However, neuroimaging research demonstrated that, although there are neural differences between genders in moral sensitivity, these do not necessarily translate into differences in the intensity of moral distress[11]. Therefore, the absence of gender-related differences in the present study may reflect the influence of modern medical curricula that emphasize integrative learning and socio-emotional competencies, which tend to mitigate gender disparities in moral and emotional formation.

Moral distress is not an inevitable consequence of hierarchical position. Factors related to the doctor–patient relationship

and the intrinsic motivation to perform good clinical work were negatively associated with distress, suggesting potential protective effects. Positive professional role models may help students cultivate resilience and moral integrity, mitigating the effects of ethically challenging situations[2].

Some items yielded unexpected results. In item S13, few students reported posting patient-related information on social media—a surprisingly low frequency given the widespread use of digital platforms. This may indicate higher ethical awareness among students or potential underreporting due to social desirability bias.

In item S20, which addressed the perception of supervisors ignoring students' suggestions, it was hypothesized that this would be a major source of moral distress given the traditional hierarchical structure of healthcare teams. However, most participants did not report feeling ignored, suggesting increased openness and dialog between supervisors and students. This may reflect evolving teaching dynamics, where preceptors encourage shared decision-making and recognize the educational value of participatory learning environments.

Similarly, in item S23, it was expected that students would report significant difficulty maintaining composure or making decisions under pressure due to limited clinical experience. Surprisingly, most participants reported feeling relatively confident in such situations. This may indicate the development of coping strategies and decision-making confidence throughout training, though self-perceived competence may not always reflect actual clinical ability.

This study presents certain limitations. Data were collected from a single institution, which may limit generalization. Although participation was voluntary and anonymous, response bias cannot be excluded. Given the sensitive topic, underreporting is possible. Online administration could be related to potential selection bias and response-condition effects. The checklist included 57 items reflecting contemporary professional dilemmas; different scenarios might yield varying responses. Moreover, no statistical correction for multiple comparisons was applied, as the analysis had an exploratory nature. Thus, some associations that appeared significant may be spurious. Finally, participants' reports of observed or experienced behaviors may not predict future professional actions. A multicenter validation is a key next step.

The findings underscore the need for medical schools to implement structured strategies that promote ethical reflection, empathic supervision, and psychological support during internship training. Incorporating formal spaces for the discussion of ethical dilemmas and encouraging open communication with mentors may reduce moral distress and strengthen professional identity formation in medical students.

CONCLUSION

More than 70% of medical interns reported experiencing some degree of moral distress, with approximately 44% describing moderate or severe stress in response to ethical or behavioral dilemmas. This prevalence suggests that moral distress is a significant reality in medical training. By quantifying experiences of moral distress, educators and institutions can better recognize the ethical tensions inherent in medical education and design targeted pedagogical interventions to promote emotional well-being, professional integrity, and resilience among trainees.

AUTHORS' CONTRIBUTIONS

GF: Data curation, Writing – original draft. **LOAF:** Data curation, Writing – original draft. **MRB:** Data curation, Writing – original draft. **RAD:** Conceptualization, Data curation, 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. Berger JT. Moral distress in medical education and training. J Gen Intern Med. 2014;29(2):395-8. https://doi.org/10.1007/s11606-013-2665-0
2. Camp M, Sadler J. Moral distress in medical student reflective writing. AJOB Empir Bioeth. 2019;10(1):70-8. https://doi.org/10.1080/23294515.2019.1570385
3. Tigard DW. Rethinking moral distress: conceptual demands for a troubling phenomenon affecting health care professionals. Med Health Care Philos. 2018;21(4):479-88. https://doi.org/10.1007/s11019-017-9819-5
4. Nagler A, Andolsek K, Rudd M, Sloane R, Musick D, Basnight L. The professionalism disconnect: do entering residents identify yet participate in unprofessional behaviors? BMC Med Educ. 2014;14:60. https://doi.org/10.1186/1472-6920-14-60
5. Song X, Ding N, Jiang N, Zhang X, Li H, Wen D. Moral distress from professionalism dilemmas and its association with self-rated professionalism behaviors among Chinese residents.

Med Teach. 2024;46(9):1210-9. https://doi.org/10.1080/0142159X.2024.2307486

6. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). 2000;25(24):3186-91. https://doi.org/10.1097/00007632-200012150-00014
7. StataCorp. Stata Statistical Software: Release 19. College Station, TX: StataCorp LLC; 2025.
8. Epstein EG, Hamric AB. Moral distress, moral residue, and the crescendo effect. J Clin Ethics. 2009;20(4):330-42. PMID: 20120853
9. Jameton A. Dilemmas of moral distress: moral responsibility and nursing practice. AWHONNS Clin Issues Perinat Womens Health Nurs. 1993;4(4):542-51. PMID: 8220368
10. Gilligan C. In a different voice: women's conceptions of self and of morality. Harvard Educ Rev. 1977;47(4):481-517. https://doi.org/10.17763/haer.47.4.g6167429416hg5l0
11. Harenski CL, Antonenko O, Shane MS, Kiehl KA. Gender differences in neural mechanisms underlying moral sensitivity. Soc Cogn Affect Neurosci. 2008;3(4):313-21. https://doi.org/10.1093/scan/nsn026

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260593

Prognostic value of the blood urea nitrogen-to-albumin ratio for 28-day mortality in septic intensive care unit patients: a large-scale retrospective cohort study using the Medical Information Mart for Intensive Care IV database

Jiatong Hou[1], Jiaqi Cheng[1], Wentao Liu[2], Yuefu Wang[1]*

SUMMARY

OBJECTIVE: Sepsis is a life-threatening condition characterized by dysregulated host responses to infection and high mortality in intensive care units. Existing prognostic models are often complex or resource-intensive. The blood urea nitrogen-to-albumin ratio, reflecting renal function, metabolic stress, and nutritional status, has emerged as a potential prognostic biomarker. The aim of this study was to evaluate the predictive value of blood urea nitrogen-to-albumin ratio for 28-day mortality in septic intensive care unit patients.

METHODS: This retrospective cohort study utilized data from the Medical Information Mart for Intensive Care IV (version 2.0) database. Adult patients meeting Sepsis-3 criteria were included. Suspected infection was defined by antibiotic administration and microbiological culture within 24 h. Demographic, clinical, and laboratory data were collected, and blood urea nitrogen-to-albumin ratio was calculated using values obtained within the first 24 h of intensive care unit admission. Receiver operating characteristic curve analysis, Kaplan-Meier survival analysis, and multivariate Cox regression were performed.

RESULTS: A total of 6,940 patients were included, with a 28-day mortality rate of 21.6%. Blood urea nitrogen-to-albumin ratio demonstrated superior predictive performance (area under the curve (AUC)=0.647, 95%CI 0.631–0.664) compared with serum urea (AUC=0.634) and albumin (AUC=0.428), and also outperformed the Sequential Organ Failure Assessment score (AUC=0.613). The optimal cutoff value was 7.6. Patients with blood urea nitrogen-to-albumin ratio ≥7.6 had significantly higher mortality than those with lower values ($p<0.001$). Elevated blood urea nitrogen-to-albumin ratio remained an independent predictor of mortality after adjustment.

CONCLUSION: Blood urea nitrogen-to-albumin ratio is a simple, objective, and readily available biomarker associated with 28-day mortality in septic intensive care unit patients and may contribute to risk stratification when used alongside other clinical parameters.

KEYWORDS: Blood urea nitrogen. Albumin. Prognosis. Sepsis. Mortality. Intensive care units.

INTRODUCTION

Sepsis, characterized by a dysregulated systemic inflammatory response to microbial invasion, remains a prevalent and life-threatening condition in the intensive care unit (ICU). Approximately one-third of critically ill patients develop sepsis, which can lead to multiple organ dysfunction and increased mortality[1]. Despite advances in organ support and sepsis management, morbidity and mortality remain unacceptably high, making prognostic evaluation a central focus in critical care medicine. Although various biomarkers and scoring systems are used to predict outcomes, many are complex, require multiple parameters, and are often costly or not readily accessible[2]. Therefore, identifying a simple, reliable, and easily obtainable prognostic indicator is of substantial clinical importance.

Blood urea nitrogen (BUN), the end product of protein metabolism synthesized in the liver and excreted by the kidneys, reflects metabolic and renal function. In sepsis, enhanced catabolism and a high risk of acute kidney injury (AKI) frequently lead to elevated BUN levels[3]. BUN is influenced by several pathophysiological factors, including renal hypoperfusion, reduced cardiac output, dehydration, and neurohumoral activation[4], and has been identified as an independent predictor of prognosis in critically ill patients[5].

Albumin, which accounts for more than 50% of plasma proteins and is synthesized by hepatocytes, plays a key role

[1]Capital Medical University, Beijing Shijitan Hospital, Department of Surgical Intensive Care Unit – Beijing, China.
[2]Capital Medical University, Beijing Shijitan Hospital, Department of Anesthesiology – Beijing, China.
*Corresponding author: wangyuefu3806@yeah.net
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on April 25, 2026. Accepted on July 29, 2026.
Scientific Editor: Roseli Nomura

in maintaining oncotic pressure and physiological homeostasis. It also serves as an indicator of nutritional status. Hypoalbuminemia is strongly associated with adverse outcomes, including increased mortality, morbidity, and prolonged ICU stay[6]. During infection and systemic inflammation, accelerated proteolysis and increased capillary permeability contribute to decreased albumin levels[5].

The BUN-to-albumin ratio (BAR), integrating both BUN and albumin, has recently emerged as a novel prognostic biomarker. It reflects both catabolic stress and inflammatory/nutritional status, and has demonstrated predictive value in various conditions, including lung cancer, gastrointestinal bleeding, and community-acquired pneumonia. Recent studies by Cai et al.[7] and Wang et al.[8] using the Medical Information Mart for Intensive Care IV (MIMIC-IV) database reported that BAR is independently associated with mortality in septic patients. However, further validation in large and independent cohorts is still required.

Accordingly, the present study aimed to evaluate the prognostic value of BAR for 28-day mortality in septic ICU patients using the MIMIC-IV database. By comparing its performance with individual markers (BUN and albumin) and the Sequential Organ Failure Assessment (SOFA) score, we sought to determine whether BAR could serve as a simple, accessible, and effective prognostic tool in clinical practice.

METHODS

Data source

Data were obtained from the MIMIC-IV (version 2.0), a large, publicly available database developed by the Massachusetts Institute of Technology. This database contains comprehensive clinical information, including demographics, vital signs, laboratory results, treatments, and outcomes for patients admitted to ICUs between 2008 and 2019. All patient data were de-identified to ensure privacy protection, and informed consent was therefore waived. The study was conducted with authorized access to the database (Certificate No. 50988851).

Study population

Adult patients (≥18 years) diagnosed with sepsis according to the Sepsis-3 criteria were included. Sepsis was defined as suspected or confirmed infection accompanied by a SOFA score ≥2 within the first 24 h of ICU admission. Suspected infection was identified based on the administration of intravenous antibiotics and microbiological culture sampling within a 24-h window, following established MIMIC-based algorithms.

The initial cohort comprised 10,578 septic patients. Exclusion criteria included: (1) ICU stay <24 h; (2) pre-existing chronic kidney disease, liver cirrhosis, or malignancy; and (3) missing BUN or albumin data within the first 24 h. After screening, 6,940 patients were included and stratified into survival and non-survival groups based on 28-day outcomes.

Data extraction and variables

Data were extracted using Structured Query Language (SQL) queries implemented through the MIMIC-IV code repository. Only the first ICU admission was considered. Extracted variables included demographic characteristics (age and sex), vital signs (heart rate, blood pressure, respiratory rate, oxygen saturation), laboratory parameters (e.g., blood cell counts, serum electrolytes, creatinine, and glucose), and comorbidities identified via International Classification of Diseases, Ninth Revision (ICD-9)/10 codes.

The SOFA score was calculated using the worst values recorded within the first 24 h of ICU admission, including PaO_2/FiO_2 ratio, platelet count, bilirubin, mean arterial pressure or vasopressor use, Glasgow Coma Scale score, and serum creatinine or urine output.

The BAR was calculated as BUN (mg/dL) divided by serum albumin (g/dL), using laboratory measurements obtained within 24 h of ICU admission.

Statistical analysis

Statistical analyses were performed using R software (version 4.2.1) and Statistical Package for the Social Sciences (SPSS) (version 26.0). The primary outcome was 28-day mortality. Continuous variables were expressed as median [interquartile range (IQR)] and compared using a t-test or Mann-Whitney U test; categorical variables were analyzed using the chi-square test. Receiver operating characteristic (ROC) curves were used to assess predictive performance, and the optimal BAR cutoff was determined by the Youden index. Internal validation was performed using 1,000 bootstrap resampling iterations to assess the stability of the area under the curve (AUC) estimates and to correct for potential overfitting. Kaplan-Meier analysis with log-rank test evaluated survival differences. Multivariate Cox regression identified independent predictors, reporting adjusted hazard ratios (HRs) with 95%CIs. Missing data were handled using complete-case analysis, and the proportion of missing values for each variable was reported. A two-tailed $p<0.05$ was considered statistically significant.

RESULTS

Extent of missing data

The proportions of missing data for each variable are summarized in Table 1. The proportion of missing values ranged from 0% (for age and gender) to 8.54% (for peripheral oxygen saturation, SO_2), with all variables showing a relatively low level of missingness. Complete-case analysis was therefore considered appropriate for the primary analysis.

Baseline demographic and clinical characteristics

In the MIMIC-IV dataset, comprising 26,941 sepsis cases as per Sepsis-3, we concentrated on patients who were ≥18 years old and admitted to the ICU for the first time. The cohort was reduced to 6,940 patients after applying the abovementioned exclusion criteria and ensuring that necessary clinical data (e.g., BUN level, albumin level, SOFA score) were available for analysis. The detailed screening process is shown in Figure 1. Table 2 delineates the baseline characteristics of both the 28-day survival group and the non-survival group. Of these, 3,056 (44%) were female, and 3,884 (56%) were male. The median age of the patient cohort was 65 years (interquartile range: 53 to 77). The overall mortality rate within 28 days was recorded at 21.6%. Comparative analysis revealed notable distinctions between the 28-day survival group and the non-survival group. The non-survival group exhibited advanced age ($p<0.001$) and presented with higher heart rate (HR), SOFA scores, and simplified acute physiology scores (SAPS II), lactate level, vasopressor use, and organ support requirement, while manifesting lower mean blood pressure (MBP), SO_2, and temperature ($p<0.001$). The median BAR value was significantly higher in non-survivors compared with survivors (14.5 [IQR: 9.2–22.0] vs. 6.8 [IQR: 3.8–11.5], $p<0.001$). Additionally, the non-survival group displayed a significantly higher incidence of AKI

Table 1. Missing data proportions for each variable.

Variable	Missing (n)	Missing (%)
BUN	1,543	5.73
Albumin	1,234	4.58
SOFA score	1,876	6.96
Heart rate	2,108	7.82
Mean blood pressure	1,987	7.38
Temperature	1,234	4.58
SO_2	2,300	8.54
Age	0	0.00
Gender	0	0.00
Hypertension	2,108	7.82
CHD	2,108	7.82
Diabetes	2,108	7.82
AKI	1,876	6.96
CRRT	1,987	7.38
Ventilation	1,543	5.73

Missing percentages were calculated based on the initial cohort of 26,941 patients. BUN: blood urea nitrogen; SOFA: Sequential Organ Failure Assessment; AKI: acute kidney injury; CRRT: continuous renal replacement therapy; CHD, coronary heart disease.

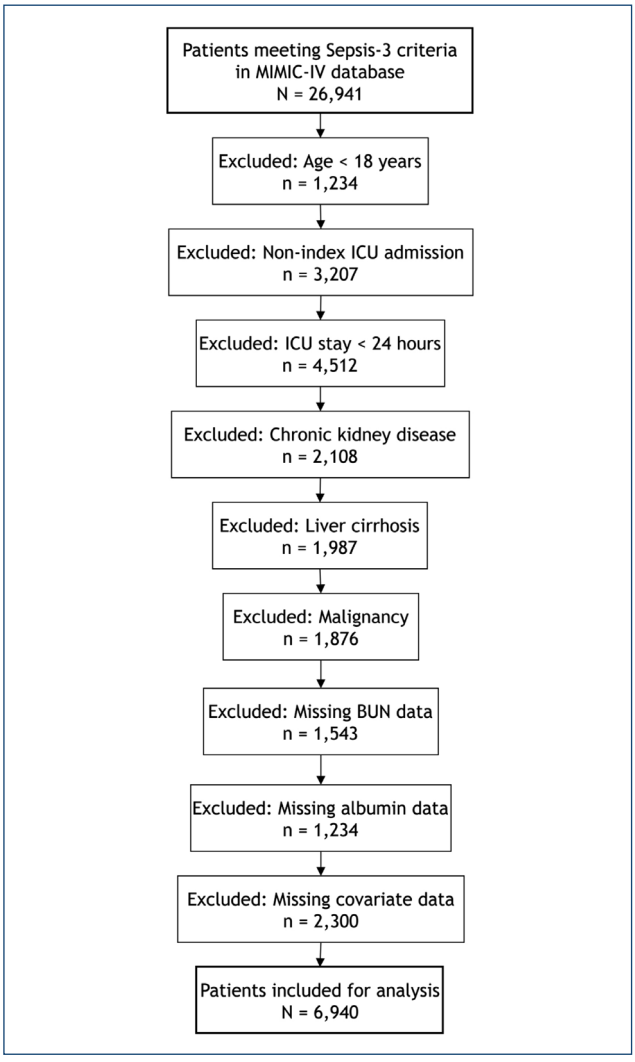


Figure 1. STROBE-compliant participant flow diagram. This flowchart illustrates the stepwise exclusion process of septic patients from the Medical Information Mart for Intensive Care IV database. Starting with 26,941 patients meeting Sepsis-3 criteria, patients were sequentially excluded due to age <18 years, non-index intensive care unit admission, intensive care unit stay <24 h, chronic kidney disease, liver cirrhosis, malignancy, missing blood urea nitrogen, missing albumin, and missing covariate data, resulting in a final cohort of 6,940 patients for analysis.

and continuous renal replacement therapy (CRRT) compared with the survival group (p<0.001). Regarding infection source, pulmonary infection was more prevalent in the non-survival group (48.0 vs. 39.1%, p<0.001).

Multivariate Cox regression analysis

After adjusting for age, sex, comorbidities [hypertension, diabetes, and coronary heart disease (CHD)], and SOFA score, SAPS II, lactate, vasopressor use, AKI, CRRT, mechanical ventilation, and infection source, multivariate Cox proportional hazards regression analysis revealed that BAR ≥7.6 was independently associated with an increased risk of 28-day mortality (adjusted HR 1.45, 95%CI 1.22–1.72, p<0.001). Other significant predictors of 28-day mortality included age (adjusted HR 1.03, 95%CI 1.02–1.04, p<0.001), SOFA score (adjusted HR 1.12, 95%CI 1.09–1.15, p<0.001), and lactate (adjusted HR 1.08,

Table 2. Comparison of baseline characteristics between 28-day survival and non-survival groups.

Variable	Total (n=6,940)	Non-survival group (n=1,500)	Survival group (n=5,440)	p
Age (year)	65 (53, 77)	72 (61, 81)	64 (52, 76)	<0.001
Gender (%) female	3,056 (44)	660 (44)	2,396 (44.1)	0.816
Gender (%) male	3,884 (56)	840 (56)	3,044 (55.9)	
Race (%) White	4,277 (61.6)	880 (58.7)	3,397 (62.5)	<0.001
Race (%) others	2,663 (38.4)	620 (41.3)	2,043 (37.5)	
HR (beat/min)	88.2 (76.3, 100.7)	95.0 (85.0, 105.0)	86.0 (75.0, 98.0)	<0.001
MBP (mmHg)	75.6 (69.6, 83.1)	70.0 (65.0, 75.0)	76.0 (70.0, 83.0)	<0.001
Temperature (℃)	36.9 (36.6, 37.3)	36.7 (36.3, 37.1)	36.9 (36.6, 37.3)	<0.001
SO_2 (%)	97.2 (95.7, 98.6)	96.0 (94.5, 97.5)	97.3 (95.8, 98.6)	<0.001
Lactate (mmol/L)	2.3 (1.5, 4.1)	4.2 (2.6, 6.8)	2.0 (1.4, 3.2)	<0.001
Hypertension (%) yes	721 (10.7)	158 (10.5)	563 (10.7)	0.885
Hypertension (%) no	6,198 (89.3)	1,342 (89.5)	4,856 (89.3)	
CHD (%) yes	368 (5.3)	100 (6.7)	268 (4.9)	0.252
CHD (%) no	6,572 (94.7)	1,400 (93.3)	5,172 (95.1)	
Diabetes (%) yes	347 (5.4)	92 (6.1)	255 (4.7)	0.914
Diabetes (%) no	6,566 (94.6)	1,408 (93.9)	5,158 (95.3)	
SOFA	3 (2, 5)	5 (3, 7)	3 (2, 4)	<0.001
SAPS II	39 (31, 50)	45 (36, 58)	37 (29, 46)	<0.001
Vasopressor use (%) yes	2,186 (31.5)	980 (65.3)	1,206 (22.2)	<0.001
Vasopressor use (%) no	4,754 (68.5)	520 (34.7)	4,234 (77.8)	
Infection source (%) pulmonary	2,846 (41.0)	720 (48.0)	2,126 (39.1)	<0.001
Infection source (%) abdominal	1,526 (22.0)	285 (19.0)	1,241 (22.8)	
Infection source (%) other	2,568 (37.0)	495 (33.0)	2,073 (38.1)	
AKI (%) yes	788 (11.4)	250 (16.7)	538 (9.9)	<0.001
AKI (%) no	6,152 (88.6)	1,250 (83.3)	4,902 (90.1)	
CRRT (%) yes	815 (11.7)	275 (18.3)	540 (9.9)	<0.001
CRRT (%) no	6,125 (88.3)	1,225 (81.7)	4,900 (90.1)	
Ventilation (%) yes	4,048 (58.3)	1,100 (73.3)	2,948 (54.2)	<0.001
Ventilation (%) no	2,892 (41.7)	400 (26.7)	2,492 (45.8)	
Organ support (%) yes	4,356 (62.8)	1,280 (85.3)	3,076 (56.5)	<0.001
Organ support (%) no	2,584 (37.2)	220 (14.7)	2,364 (43.5)	

HR: heart rate; MBP: mean blood pressure; SO_2: blood oxygen saturation; SAPS II: Simplified Acute Physiology Score II; AKI: acute kidney injury; CRRT: continuous renal replacement therapy; SAPS II: simplified acute physiology scores; CHD, coronary heart disease. Units—HR (beats/min), MBP (mmHg), temperature (°C), SO_2 (%), age (years).

95%CI 1.04–1.12, p<0.001), and vasopressor use (adjusted HR 1.35, 95%CI 1.15–1.58, p<0.001). When BAR was analyzed as a continuous variable, each one-unit increase in BAR was associated with a 3.2% increase in the risk of 28-day mortality after adjusting for all confounders (adjusted HR 1.032, 95%CI 1.018–1.046, p<0.001). These findings indicate that BAR, along with other clinical factors, plays a critical role in predicting 28-day mortality.

Receiver operating characteristic curve analysis

ROC curves for the prediction of all-cause mortality in septic patients were plotted for BAR, BUN, albumin, and SOFA, as depicted in Figure 2. The AUC for BAR was notably higher at 0.647 compared with BUN (AUC, 0.634), albumin (AUC, 0.428), and SOFA (AUC, 0.613). This discrepancy underscores the superior predictive efficacy of BAR. Moreover, the optimal critical value for BAR was identified as 7.6, with a sensitivity of 0.683 and a specificity of 0.55. To assess the stability of the model, internal validation was performed using 1,000 bootstrap resampling iterations. The bootstrap-corrected AUC for BAR was 0.641 (95%CI 0.625–0.658), consistent with the original estimate, indicating good internal validity of the predictive model.

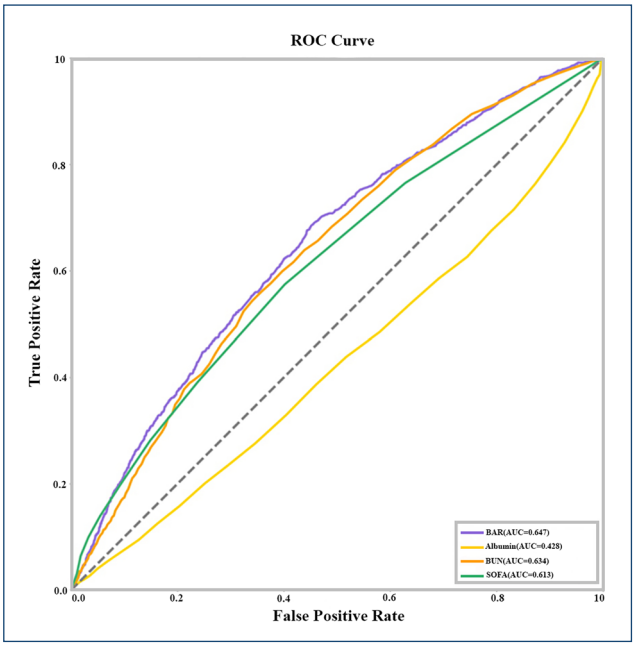


Figure 2. Receiver operating characteristic curves of the blood urea nitrogen-to-albumin ratio, blood urea nitrogen, serum albumin, and the Sequential Organ Failure Assessment score for predicting 28-day mortality in septic intensive care unit patients. The receiver operating characteristic curves compare the predictive performance of blood urea nitrogen-to-albumin ratio (purple solid line), blood urea nitrogen (orange solid line), Sequential Organ Failure Assessment score (green solid line), and serum albumin (yellow solid line). Blood urea nitrogen-to-albumin ratio exhibited the highest area under the curve (AUC=0.647), demonstrating its superior prognostic accuracy for 28-day mortality over the individual markers and the conventional multi-parameter scoring system.

Kaplan-Meier survival curve

According to the best critical value, patients were divided into high BAR group (BAR ≥7.6, n=3,436) and low BAR group (BAR<7.6, n=3,504). Among these, 1,250 patients (36.4%) in the high BAR group and 250 patients (7.1%) in the low BAR group died within 28 days. The Kaplan-Meier survival curve is illustrated in Figure 3. The results indicated that patients in the high BAR group had a significantly lower survival rate than those in the low BAR group (log-rank p<0.001).

DISCUSSION

The present study, based on data from the MIMIC-IV (version 2.0) database, included 6,940 septic patients and demonstrated a 28-day mortality rate of 18.7%. The BAR showed superior predictive performance for mortality (AUC=0.647) compared with serum urea (AUC=0.634) and serum albumin (AUC=0.428), and also outperformed the SOFA score (AUC=0.613). The optimal cutoff value for BAR was identified as 7.6, and Kaplan-Meier analysis revealed significantly

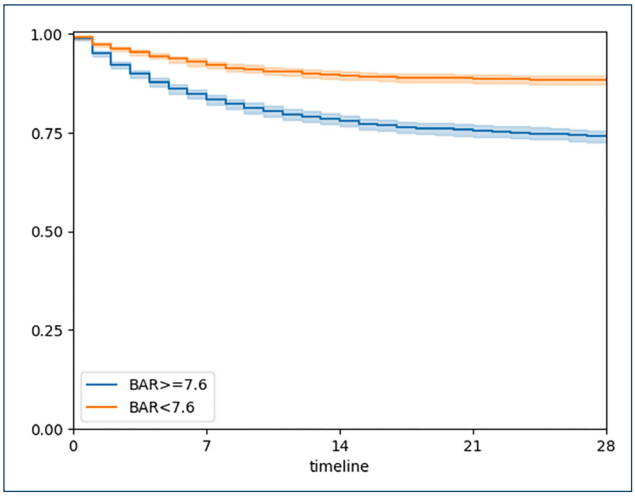


Figure 3. Kaplan-Meier survival analysis curves for all-cause mortality within 28 days of hospital admission. This figure presents the Kaplan-Meier survival analysis based on the optimal critical value of blood urea nitrogen-to-albumin ratio (7.6). Patients are categorized into high blood urea nitrogen-to-albumin ratio (BAR ≥7.6) and low blood urea nitrogen-to-albumin ratio (BAR <7.6) groups, and the survival curves depict the cumulative survival rates over the 28-day period. The results show a significantly higher survival rate in the high blood urea nitrogen-to-albumin ratio group compared with the low blood urea nitrogen-to-albumin ratio group (p<0.001), emphasizing the prognostic value of blood urea nitrogen-to-albumin ratio in predicting short-term mortality.

higher mortality among patients with BAR ≥7.6. These findings suggest that BAR may serve as a useful adjunctive marker for risk assessment in septic ICU patients, though its modest discriminative ability indicates that it should be interpreted in conjunction with other clinical parameters.

Sepsis remains a major cause of morbidity and mortality in critically ill patients, characterized by a dysregulated host response to infection leading to life-threatening organ dysfunction[8]. Despite continuous improvements in supportive care and therapeutic strategies, the burden of sepsis remains substantial. Therefore, identifying simple and reliable prognostic biomarkers is essential for early risk stratification and timely clinical intervention.

BUN is widely recognized as an indicator of renal function, metabolic status, and protein catabolism[9]. In critically ill patients, systemic stress responses promote protein breakdown, leading to increased BUN production. In addition, enhanced renal tubular reabsorption further contributes to elevated BUN levels[10]. Previous studies have demonstrated that elevated BUN levels at ICU admission, as well as peak BUN levels during hospitalization, are independently associated with increased mortality risk[4,11]. These changes are primarily attributed to intensified protein catabolism and dehydration in critically ill patients[11].

Albumin, which constitutes more than 50% of plasma proteins, plays a critical role in maintaining oncotic pressure, modulating inflammatory responses, and protecting against ischemia-reperfusion injury[12]. In sepsis, systemic inflammation triggers cytokine release, leading to increased vascular permeability, capillary leakage, and tissue hypoperfusion, all of which contribute to hypoalbuminemia. Reduced albumin levels are consistently associated with poor clinical outcomes and have been identified as independent predictors of mortality in patients with sepsis and septic shock.

Although the SOFA score is widely used for assessing organ dysfunction, its application may be limited in certain clinical settings due to its reliance on multiple parameters. In contrast, BAR integrates two routinely available laboratory indices—BUN and albumin—into a single parameter that simultaneously reflects renal function, metabolic stress, inflammation, and nutritional status. Elevated BUN reflects renal hypoperfusion and catabolic stress, while hypoalbuminemia indicates systemic inflammation and poor nutritional reserves. By combining these complementary dimensions, BAR provides a more comprehensive representation of the pathophysiological alterations in sepsis, which may explain its superior predictive performance compared with individual markers and the SOFA score. Moreover, BUN and albumin measurements are objective, cost-effective, and easily obtainable, enhancing the clinical applicability of BAR.

The present findings are consistent with and extend previous studies by Cai et al.[7] and Wang et al.[8], both of which demonstrated that BAR is independently associated with mortality in septic patients using the MIMIC-IV database. Recent studies have further reported the prognostic value of BAR in specific sepsis subpopulations. Han et al. demonstrated that BAR was positively correlated with 28-day mortality in patients with sepsis-induced AKI, with an AUC of 0.639 and an HR of 1.029 per unit increase[13]. Li et al. reported that in patients with sepsis-induced coagulopathy, each 10-unit increase in BAR was associated with a 30% higher hazard of 30-day mortality (adjusted HR 1.18, 95%CI 1.12–1.24)[14]. Compared with prior research, this study offers several methodological strengths. It utilized the updated MIMIC-IV v2.0 database, applied the Sepsis-3 criteria for case identification, and employed rigorous exclusion criteria to minimize confounding factors, such as chronic kidney disease, liver cirrhosis, and malignancy. Furthermore, sepsis was identified using a combination of SOFA scores and standardized definitions of suspected infection, rather than relying solely on administrative coding. Clinically, this study confirmed that BAR outperforms both its individual components and the SOFA score in predicting 28-day mortality in a large cohort of septic ICU patients, thereby enhancing the robustness and generalizability of the findings. Beyond these prior reports, our study provides several novel contributions. First, we evaluated BAR in a broad, unselected septic ICU population (n=6,940), whereas previous studies focused on specific subpopulations. Second, we systematically compared BAR's predictive performance against both its individual components (BUN and albumin) and the SOFA score. Third, we established a clinically applicable optimal cutoff value (7.6) and performed internal validation. Fourth, we integrated multiple clinically relevant confounders including lactate, vasopressor use, AKI, CRRT, mechanical ventilation, and infection source into the multivariable model, providing a more robust adjustment for disease severity.

Nevertheless, several limitations should be acknowledged. First, the relatively modest AUC value (0.647) suggests that BAR, although informative, is not a definitive predictor of mortality and may require integration with other biomarkers or clinical parameters to improve predictive accuracy. Second, the retrospective design introduces potential residual confounding and limits causal inference. Third, BAR was calculated based on a single measurement within the first 24 h of ICU admission, which may not fully capture the dynamic progression of sepsis. Serial monitoring of BAR may provide additional prognostic

insights. Fourth, BAR measurements may be influenced by several clinical factors. Exogenous albumin administration, commonly used for volume resuscitation or albumin replacement in septic patients, may directly elevate serum albumin levels and consequently lower BAR values, potentially underestimating mortality risk. Fluid resuscitation, while essential for hemodynamic support, may dilute both BUN and albumin concentrations, introducing variability in BAR measurements that does not necessarily reflect true pathophysiological changes. Moreover, AKI is prevalent in sepsis and directly affects BUN levels through impaired renal excretion, potentially confounding the association between BAR and mortality. While we adjusted for AKI in the multivariable Cox regression model, residual confounding from these dynamic factors cannot be fully excluded. Finally, while complete-case analysis was used to handle missing data, the exclusion of patients with missing data may have introduced selection bias. However, given the low proportion of missing values across all variables, the potential impact of this bias is likely minimal. Future prospective, multicenter studies are warranted to validate these findings across diverse populations and to explore the value of dynamic BAR monitoring in clinical practice.

CONCLUSION

This study demonstrated that the BAR is a simple, objective, and clinically accessible biomarker associated with 28-day mortality in septic ICU patients. Using data from the MIMIC-IV database, BAR exhibited modest predictive performance, which was superior to serum urea, serum albumin, and the SOFA score, with an optimal cutoff value of 7.6. Patients with BAR ≥7.6 had significantly higher mortality risk. These findings suggest that BAR may contribute to early risk stratification when used alongside other clinical parameters, but its modest discriminative ability underscores the need for caution in clinical application. Further prospective studies are needed to validate these results and to investigate whether combining BAR with other biomarkers can further enhance prognostic accuracy.

ETHICS STATEMENT

Ethical review and approval were waived for this study because all data were obtained from a publicly available, de-identified database (MIMIC-IV), and no identifiable human participants were involved. Therefore, written informed consent for participation was not required in accordance with national legislation and institutional requirements.

AUTHORS' CONTRIBUTIONS

JH: Conceptualization, Writing – original draft. **JC:** Data curation, Formal Analysis, Investigation, Validation. **WL:** Data curation, Formal Analysis, Investigation, Validation. **YW:** Conceptualization, 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. Hollenberg SM, Singer M. Pathophysiology of sepsis-induced cardiomyopathy. Nat Rev Cardiol. 2021;18(6):424-34. https://doi.org/10.1038/s41569-020-00492-2
2. Wu Q, Xiao Z, Pu Y, Zhou J, Wang D, Huang Z, et al. TnI and IL-18 levels are associated with prognosis of sepsis. Postgrad Med J. 2019;95(1123):240-4. https://doi.org/10.1136/postgradmedj-2018-136371
3. Sharma K, Mogensen KM, Robinson MK. Pathophysiology of critical illness and role of nutrition. nutrition in clinical practice: official publication of the American Society for Parenteral and Enteral Nutrition. 2019;34(1):12-22. https://doi.org/10.1002/ncp.10232
4. Dundar ZD, Kucukceran K, Ayranci MK. Blood urea nitrogen to albumin ratio is a predictor of in-hospital mortality in older emergency department patients. Am J Emerg Med. 2021;46:349-54. https://doi.org/10.1016/j.ajem.2020.10.008
5. Zeng Z, Ke X, Gong S, Huang X, Liu Q, Huang X, et al. Blood urea nitrogen to serum albumin ratio: a good predictor of in-hospital and 90-day all-cause mortality in patients with acute exacerbations of chronic obstructive pulmonary disease. Bmc Pulm Med. 2022;22(1):476. https://doi.org/10.1186/s12890-022-02258-7
6. Pereira AG, Costa NA, Gut AL, Azevedo PS, Tanni SE, Mamede Zornoff LA, et al. Urea to albumin ratio is a predictor of mortality in patients with septic shock. Clin Nutr ESPEN. 2021;42:361-5. https://doi.org/10.1016/j.clnesp.2021.01.007
7. Cai S, Wang Q, Chen C, Guo C, Zheng L, Yuan M. Association between blood urea nitrogen to serum albumin ratio and in-hospital mortality of patients with sepsis in intensive care: a retrospective analysis of the fourth-generation Medical Information Mart for Intensive Care database. Front Nutr. 2022;9:967332. https://doi.org/10.3389/fnut.2022.967332
8. Wang Y, Gao S, Hong L, Hou T, Liu H, Li M, et al. Prognostic impact of blood urea nitrogen to albumin ratio on patients with sepsis: a retrospective cohort study. Sci Rep-Uk. 2023;13(1):10013. https://doi.org/10.1038/s41598-023-37127-8
9. Min J, Lu J, Zhong L, Yuan M, Xu Y. The correlation study between blood urea nitrogen to serum albumin ratio and prognosis of patients with sepsis during hospitalization. BMC Anesthesiol. 2022;22(1):404. https://doi.org/10.1186/s12871-022-01947-4

10. Hung K, Li Y, Huang Y, Liu P, Hsu C, Ho C, et al. Efficacy of blood urea nitrogen-to-albumin ratio for predicting prognostic outcomes of inpatients with COVID-19: a meta-analysis. Medicine. 2023;102(7):e33007. https://doi.org/10.1097/MD.0000000000033007
11. Fang J, Xu B. Blood urea nitrogen to serum albumin ratio independently predicts mortality in critically Ill patients with acute pulmonary embolism. Clin Appl Thromb Hemost. 2021;27:10760296211010241. https://doi.org/10.1177/10760296211010241
12. Etzrodt V, Idowu TO, Schenk H, Seeliger B, Prasse A, Thamm K, et al. Role of endothelial microRNA 155 on capillary leakage in systemic inflammation. Crit Care (London, England). 2021;25(1):76. https://doi.org/10.1186/s13054-021-03500-0
13. Han K, Tao YX, Wang JH, Lu JS. Prognostic value of blood urea nitrogen to albumin ratio in septic patients with acute kidney injury-a retrospective study based on MIMIC database. Front Med (Lausanne). 2025;12:1510919. https://doi.org/10.3389/fmed.2025.1510919
14. Li M, Wang HB, Wang H, Xu ZJ. Association of blood urea nitrogen to serum albumin ratio with mortality in ICU patients with sepsis-induced coagulopathy: a retrospective cohort study from MIMIC-IV database. Medicine (Baltimore). 2026;105(15):e48296. https://doi.org/10.1097/MD.0000000000048296

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260627

Computed tomography of the occipital bone for sex estimation in Turks

Esra Arslan[1], Veysel Osman Soğukpınar[2]*, Melahat Kul[3], Yaşar Bilge[4]

SUMMARY

OBJECTIVE: Sex estimation is a fundamental component of forensic identification when DNA analysis is not feasible. Because cranial morphology and sexual dimorphism vary among populations, population-specific standards are required. However, occipital bone morphometric data for Turkish adults remain limited. Therefore, the aim of this study was to evaluate the reliability of occipital bone measurements obtained from cranial computed tomography for sex estimation in Turkish adults.

METHODS: Three-dimensional cranial computed tomography images of 384 individuals (192 females, 192 males), aged 18–65, obtained between 01/01/2010 and 31/12/2021, were retrospectively analyzed. Nine anatomical distances on the occipital bone were measured, and their discriminative roles in sex estimation were assessed using logistic regression analysis. Additionally, inter-observer agreement was evaluated by comparing measurements from two independent observers.

RESULTS: The combined analysis of all measurements yielded a sex prediction accuracy of 82.8% for females and 80.2% for males. Foramen magnum width, occipital condyle length, occipital condyle width, and external occipital protuberance depth were identified as the most discriminative variables. A model combining these four variables achieved classification accuracies of 83.3% for females and 79.2% for males, with an overall accuracy of 81.3%. The full model including all nine measurements showed an overall accuracy of 81.5%. High intra- and inter-observer agreement was observed.

CONCLUSION: Computed tomography-based morphometric analysis of the occipital bone is a reliable method for sex estimation in forensic medicine. These results support the development of population-specific reference data for the Turkish population.

KEYWORDS: Sex determination by skeleton. Forensic anthropology. Tomography. Skull. Anthropometry.

INTRODUCTION

Identity encompasses the biological and sociological characteristics that distinguish an individual; determining these characteristics constitutes identification. In forensic and social contexts, identification applies to both living and deceased persons and is categorized as forensic identity and medical identity. Forensic identity includes demographic data (sex, date and place of birth, parents' names) and official documentation such as identity cards, driver's licenses, and passports. Medical identity refers to physical and anatomical features including stature, body weight, age, pigmentation traits, facial morphology, scars, surgical marks, tattoos, birthmarks, skeletal anomalies, and dental characteristics[1].

In forensic medicine, sex estimation is a fundamental component of constructing the biological profile of unidentified individuals[2]. Accurate sex determination narrows missing person searches and guides assessment of age, ancestry, and stature. Although DNA profiling provides high accuracy, it may be limited by cost, time requirements, or lack of comparative samples. In such cases, anthropological and morphometric methods serve as important alternatives, particularly when rapid results are required[3].

Sexual dimorphism of the human skeleton is well established, and the cranium is one of the most extensively studied regions for sex estimation. Shim et al.[4] reported approximately 93% accuracy using three-dimensional cranial metric analysis. Recent research trends indicate that metric methods are increasingly prioritized over non-metric (morphological) approaches in forensic sex estimation due to their quantifiable and replicable nature, which tends to reduce observer-dependent variability[5]. In fragmented remains, the pelvis and cranium are generally regarded as the most informative skeletal elements for sex estimation, although other bones have also been investigated[3,6,7]. Advances in computed tomography (CT)

[1]Ministry of Justice, Council of Forensic Medicine, The Group Presidency of Forensic Medicine – Ankara, Türkiye.
[2]Ministry of Justice, Council of Forensic Medicine, The Group Presidency of Forensic Medicine – Van, Türkiye.
[3]Ankara University, Faculty of Medicine, Department of Internal Medicine, Division of Radiology – Ankara, Türkiye.
[4]Ankara University, Faculty of Medicine, Department of Internal Medicine, Division of Forensic Medicine – Ankara, Türkiye.
*Corresponding author: v.o.s23@hotmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on June 02, 2026. Accepted on August 08, 2026.
Scientific Editor: José Maria Soares Júnior

and magnetic resonance imaging (MRI) enable rapid, objective, and three-dimensional evaluation of skeletal structures. Moreover, cranial CT allows detailed morphometric assessment without the need for direct skeletal examination and has become increasingly available in both clinical and forensic settings. Since cranial morphology and the degree of sexual dimorphism differ among populations because of genetic, environmental, and geographical factors, population-specific reference standards are essential for accurate sex estimation[8]. However, studies evaluating occipital bone morphometry in the Turkish population remain limited. Therefore, the present study aimed to investigate sexual dimorphism in linear measurements of the occipital bone using cranial CT and to evaluate their reliability for sex estimation in Turkish adults.

METHODS

Study design

This descriptive cross-sectional study assessed the applicability of occipital bone linear measurements on CT images for sex estimation in individuals aged 18–65 years.

Sample size was determined based on planned statistical analyses and literature-reported effect sizes (0.15, 0.30, 0.40, and 0.50) with 95% power, requiring a minimum of 384 participants. Accordingly, cranial CT scans of 384 individuals (192 females and 192 males) aged 18–65 years, obtained at the Department of Radiology, Ankara University Faculty of Medicine between 01/01/2010 and 31/12/2021, were included. Participants were evenly distributed across age subgroups (18–24, 25–30, 31–39, 40–49, 50–59, and 60–65).

Exclusion criteria comprised prior cranial base surgery, trauma-related deformity, and tumors, masses, or hydrocephalus involving the cranial base.

Demographic and clinical data were retrieved from the AVICENNA hospital information management system. All CT images were evaluated collaboratively by radiology and forensic medicine specialists.

Radiological evaluation

All CT images were analyzed using VITREA advanced visualization software (v7.5.0.510, Toshiba) with multiplanar and three-dimensional reconstructions. Examinations were performed using Somatom Emotion (Siemens, 16-slice) and Prime Aquillon (Toshiba, 80-slice) scanners. CT acquisition parameters included a slice thickness of 0.5–0.625 mm, tube current of 140–300 mA, tube voltage of 120–140 kVp, and pitch values of 0.348–0.637.

Linear measurements of the occipital bone were obtained in millimeters (mm) and recorded in centimeters (cm). These measurements included: external occipital protuberance depth (EOPD) (perpendicular distance from the most prominent point to the bone base), foramen magnum width (FMW) (maximum width measured perpendicular to the midline), foramen magnum length (FML) (maximum distance between basion and opisthion), inion-opisthion length (IOL) (external measurement), lambda-inion length (LIL) (external measurement), lambda-opisthion length (LOL) (external measurement), occipital condyle width (OCW) (maximum width perpendicular to the long axis of the condyle), occipital condyle length (OCL) (maximum length along the long axis of the condyle), and interasterion distance (IAD) (external distance between the right and left asterion). Representative three-dimensional and axial cranial CT images illustrating the anatomical landmarks and morphometric measurements used in the present study are shown in Figure 1.

Statistical analysis

Statistical analyses were performed using statistical package for the social sciences (SPSS) version 26.0 (Armonk, New York, USA); $p<0.05$ was considered statistically significant.

Continuous variables were expressed as mean±standard deviation and categorical variables as number and percentage. Distribution was assessed using histograms, probability plots, and the D'Agostino-Pearson, Kolmogorov-Smirnov, and Shapiro-Wilk tests. Outliers in non-normally distributed data were removed, and Winsorization was applied when required. Homogeneity of variances was evaluated using the Levene test.

Between-sex comparisons were performed using the independent t-test or Welch t-test; the Kruskal-Wallis test was used for non-normal distributions, and one-way analysis of variance (ANOVA) for normally distributed comparisons involving more than two groups.

Intra- and inter-observer agreement were assessed using the Paired t-test and Wilcoxon Paired test. The chi-square test and Binary Logistic Regression were used to determine diagnostic performance in sex discrimination, and model fit was evaluated with the Omnibus test.

Ethics committee approval

Approval for this study was obtained from the Ankara University Faculty of Medicine Human Research Ethics Committee (decision number İ04-275-23, dated 10/05/2023), and the principles of the Declaration of Helsinki were taken into consideration.

RESULTS

A total of 384 individuals (192 males, 192 females; 18–65 years) were included.

Except for IOL (p=0.05), all occipital measurements (LOL, LIL, OCL, OCW, EOPD, IAD, FML, and FMW) showed significant sex differences (p<0.001). Welch's t-test was applied to LOL, EOPD, and IAD due to variance heterogeneity; the independent samples t-test was used for the remaining variables. For all significant parameters, females exhibited lower mean values than males (Table 1).

Binary Logistic Regression including all variables (IOL, LOL, LIL, OCL, OCW, EOPD, IAD, FML, and FMW) demonstrated overall model significance (Omnibus p<0.001). The model correctly classified 82.8% (n=159) of females and 80.2% (n=154) of males, with an overall accuracy of 81.5% (n=313). OCL, OCW, EOPD, and FMW were significant predictors; increasing values were associated with male sex (Table 2).

Seven additional measurement combinations were tested. All models were statistically significant. The highest accuracy (81.3%) was obtained with the OCL-OCW-FMW-EOPD model, whereas the lowest (61.2%) was observed for LOL-LIL-IOL. In all models, higher values of significant variables were positively associated with male sex (Table 3).

Inter- and intra-observer reliability was assessed in 48 cases evaluated independently by a forensic medicine faculty member and a radiology specialist. After normality testing, the Paired t-test (parametric variables) and Wilcoxon Paired test (IOL) were applied. No significant inter-observer differences were detected for any measurement (p>0.05), indicating high measurement consistency.

DISCUSSION

In forensic practice, identification is the primary step for both living and deceased individuals, and sex estimation constitutes its foundation. Accurate sex determination directly affects the reliability of the remaining components of the biological profile. Although DNA analysis remains the gold standard through X and Y chromosome differentiation[9], morphological and metric methods are indispensable when DNA is unavailable, particularly in degraded or limited samples[10]. Therefore, standardized anthropological methods applicable to skeletal remains should be maintained as complementary tools in forensic identification.

Bone tissue may be the only available source of genetic material in certain cases due to its relative capacity to preserve DNA[11]. However, environmental and physical factors may

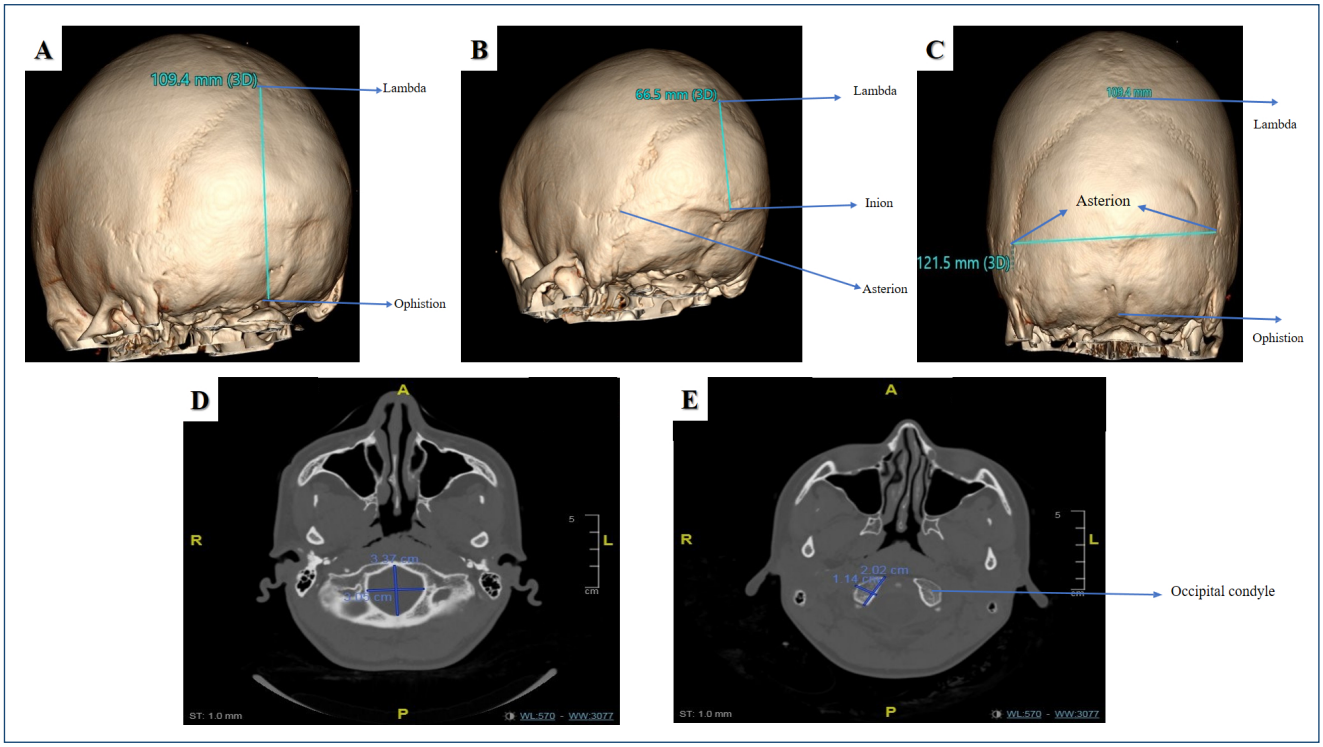


Figure 1. Representative three-dimensional and axial cranial computed tomography images illustrating the anatomical landmarks and morphometric measurements used in this study. **(A)** Lambda–opisthion length. **(B)** Lambda–inion length. **(C)** Interasterionic distance. **(D)** Foramen magnum length and foramen magnum width. (E) Occipital condyle length and occipital condyle width. A: anterior; P: posterior; R: right; L: left.

compromise DNA integrity. Acidic soil conditions, chemical exposure, and high temperatures—such as in fire-related events—can cause denaturation and severe degradation[12]. Consequently, DNA extraction from bone is technically complex and does not consistently yield usable results. These limitations reinforce the necessity of morphometric approaches when genetic analysis is restricted.

In this study, the diagnostic value of the occipital bone was evaluated in forensic scenarios involving limited material, degraded remains, or time constraints. In mass fatalities, commingled remains, or fragmented skeletons, the pelvis, long bones, and cranium are primary structures for sex estimation[13]. Using nine linear measurements from anatomical landmarks of the occipital bone—one of the most durable cranial regions—an overall accuracy of 81.5% was achieved. Previous studies based solely on foramen magnum or occipital measurements reported accuracies ranging from 66.5 to 88%[7,14-17], while models incorporating additional cranial parameters reached 87%[18].

Table 1. Cranial (occipital bone) measurement values by sex and statistical comparison results.

Measurements (parameters)	Levene test of variance (p)	Levene variance homogeneity (+/-)	Test type	t-value	p-value	Average difference (cm)	Average value for women (cm)	Average value for men (cm)
IOL	0.88	+	Independent t-test	1.95	0.05	0.1243	4.74	4.87
LOL	<0.05	-	Welch t-test	5.82	<0.001	0.4174	9.91	10.33
LIL	0.50	+	Independent t-test	5.69	<0.001	0.4900	6.41	6.90
OCL	0.82	+	Independent t-test	10.50	<0.001	0.2084	1.72	1.92
OCW	0.12	+	Independent t-test	7.79	<0.001	0.1027	0.87	0.97
EOPD	0.02	-	Welch t-test	10.77	<0.001	0.3435	1.07	1.42
IAD	0.02	-	Welch t-test	9.27	<0.001	0.5016	11.09	11.59
FML	0.43	+	Independent t-test	9.62	<0.001	0.2670	3.33	3.59
FMW	0.23	+	Independent t-test	8.46	<0.001	0.2081	2.88	3.08

OCL: occipital condyle length; OCW, occipital condyle width; EOPD, external occipital protuberance depth; LOL, lambda-opisthion length; LIL, lambda-inion length; IOL, inion-opisthion length; FML, foramen magnum length; FMW, foramen magnum width; IAD, interasterion distance.

Table 2. Sex estimation based on cranial (occipital bone) measurements and binary logistic regression results.

Independent variable parameters (measurements)	Regression coefficient	Standard error	Wald statistic (chi-square)	Degree of freedom	Significance level (p)	Probability ratio
LOL	0.079	0.133	0.352	1	0.553	1.082
LIL	0.051	0.039	1.686	1	0.194	1.052
IOL	-0.037	0.063	0.351	1	0.553	0.964
OCL	3.332	0.823	16.377	1	<0.001	28.005
OCW	3.510	1.230	8.143	1	0.004	33.441
EOPD	3.567	0.527	45.740	1	<0.001	35.418
IAD	-0.026	0.071	0.138	1	0.711	0.974
FML	-0.016	0.084	0.037	1	0.848	0.984
FMW	4.199	0.660	40.422	1	<0.001	66.614
Fixed	-26.740	3.127	73.119	1	<0.001	0.000

OCL: occipital condyle length; OCW, occipital condyle width; EOPD, external occipital protuberance depth; LOL, lambda-opisthion length; LIL, lambda-onion length; IOL, inion-opisthion length; FML, foramen magnum length; FMW, foramen magnum width; IAD, interasterion distance.

Table 3. Logistic regression analysis of cranial (occipital bone) metric parameters (measurements) in sex differentiation.

Combination models	Omnibus test (p)	Correct classification "%" (female/male/overall average)	Correct classification "n" (female/male/total)	Significant parameters (measurements)	Comment
OCL, OCW, EOPD	<0.001	79.7%/76%/77.9%	153/146/299	OCL, OCW, EOPD	An increase in all measurements indicates a transition toward the male sex.
LOL, LIL, IOL	0.004	64.1%/58.3%/61.2%	123/112/235	Only LOL	An increase in the LOL measurement indicates a transition toward the male sex.
FML, FMW	0.004	67.2%/67.7%/67.4%	129/130/259	Only FMW	An increase in the FMW measurement indicates a transition toward the male sex.
OCL, OCW, IAD	<0.001	74%/71.4%/72.7%	142/137/279	OCL, OCW	An increase in the OCL and OCW measurements indicates a transition toward the male sex.
LOL, IAD	<0.001	72.4%/66.7%/69.5%	139/128/267	LOL, IAD	An increase in both measurements indicates a transition toward the male sex.
OCL, OCW, FML, FMW	<0.001	75.5%/71.4%/73.4%	145/137/282	OCL, OCW, FMW	The FML measurement is not significant; increases in the others indicate a transition toward the male sex.
OCL, OCW, FMW, EOPD	<0.001	83.3%/79.2%/81.3%	160/152/312	OCL, OCW, FMW, EOPD	An increase in all measurements indicates a transition toward the male sex.

OCL: occipital condyle length; OCW: occipital condyle width; EOPD, external occipital protuberance depth; LOL, lambda-opisthion length; LIL, lambda-inion length; IOL, inion-opisthion length; FML, foramen magnum length; FMW, foramen magnum width; IAD, interasterion distance.

These findings collectively support the diagnostic relevance of occipital bone metrics in sex estimation and their integration into routine forensic protocols.

Kirisattayakul et al.[19] demonstrated sexual dimorphism of the external occipital protuberance (EOP) in a Thai population using cranial CT. In the present study, EOP was assessed metrically through EOPD (1.07 cm in females; 1.42 cm in males), confirming dimorphism while emphasizing the objectivity of quantitative evaluation. The growing application of three-dimensional imaging in forensic anthropology enhances measurement precision, reproducibility, and efficiency, enabling retrospective analyses and rapid identification.

Except for IOL, all occipital parameters showed significant sexual dimorphism. Similar findings were reported by Ekizoglu et al.[18] in a Turkish cohort (87% accuracy), Kelley and Tallman[20] (81–87% accuracy), and Meral et al.[21] (up to 88% accuracy). Compared with the 81.5% overall and 81.3% best-model accuracies obtained here, occipital measurements demonstrate performance comparable to established cranial metric methods. High intra- and inter-observer agreement further confirms methodological reliability.

Foramen magnum-based sex estimation shows population variability. Gapert et al.[15] and Suazo et al.[16] identified FMW as the most discriminative parameter, whereas Raghavendra Babu et al.[7] reported higher dimorphism for FML. In the present study, FMW remained statistically significant, while FML showed lower predictive value (overall 67.4% when analyzed together). These inconsistencies likely reflect geographic, genetic, and methodological differences. Thus, while foramen magnum metrics may be useful in specific populations, standardized multicenter studies are necessary to enhance generalizability.

Population variability is also evident in other cranial regions. Zheng et al.[22] reported sex estimation accuracies of 62.4% (Bosnian), 75.2% (Chinese), and 78.4% (Nepalese) using frontal sinus morphometry, while a Chinese study reported 76.6% accuracy[23]. These findings emphasize the influence of genetic and environmental factors on cranial dimorphism and highlight the need for population-specific reference data. The present study contributes such data for Turkish adults.

Occipital condyle morphometry has demonstrated moderate-to-high diagnostic performance. Gapert et al.[14] reported accuracies of 69.2–76.7% using condylar measurements, while a systematic review by Tsiouris et al.[24] confirmed the value of CT-based three-dimensional evaluation of this region. In the present study, OCL and OCW were significant predictors and achieved 81.3% accuracy when combined with FMW and

EOPD. Similarly, Hoover and Thomas[17] reported 71% accuracy using combined occipital parameters, including EOPD and FMW, reinforcing their consistent discriminative value.

The relative resistance of occipital structures to trauma, burning, and decomposition further enhances forensic applicability. Additionally, the absence of significant measurement differences between observers in this study confirms high reproducibility. MRI-based studies have likewise demonstrated strong intra- and inter-observer agreement for cranial base parameters[25], supporting the reliability of advanced imaging–based morphometry.

From a practical perspective, the proposed CT-based morphometric approach may serve as a valuable adjunct in forensic identification, particularly when highly sexually dimorphic skeletal elements, such as the pelvis, are unavailable. Because cranial CT examinations are routinely performed in many clinical settings, retrospective imaging datasets may provide an accessible source for developing and validating population-specific forensic standards. Although the proposed model should not replace comprehensive anthropological or genetic analyses, it may contribute to the biological profile as a complementary method when conventional approaches are limited.

The present study has several strengths. First, it included a relatively large and sex-balanced sample of Turkish adults. Second, all measurements were obtained from three-dimensional CT images using standardized anatomical landmarks, thereby reducing measurement variability. Third, intra- and inter-observer reliability analyses demonstrated high reproducibility of the measurements. Finally, the evaluation of multiple occipital parameters within the same predictive model provides comprehensive reference data for forensic applications in the Turkish population.

This study also has several limitations. First, its retrospective single-center design may limit the generalizability of the findings. Second, the study population consisted only of Turkish adults and may not fully represent all regional and ethnic variations within the country. Third, no external validation cohort was available to evaluate the predictive model independently. Therefore, future multicenter studies including larger and more diverse populations are needed to externally validate the proposed model and improve its generalizability.

Overall, the findings of the present study indicate that CT-based occipital bone morphometry is a reproducible, objective, and non-invasive approach that may assist sex estimation in Turkish adults. Although predictive performance remains population-dependent, the proposed model provides valuable complementary information for forensic identification. Future multicenter studies with external validation cohorts are warranted to further improve the robustness and generalizability of these population-specific reference data.

CONCLUSION

CT-based occipital morphometry represents a reliable, objective, and non-invasive complementary tool for sex estimation in Turkish adults. FMW, OCL, OCW, and EOPD demonstrated significant sexual dimorphism, with the four-parameter logistic regression model achieving an accuracy of 81.3%, suggesting that a reduced model may be sufficient for practical forensic applications. The postmortem durability of the occipital region and the high inter-observer agreement further support its forensic utility. These findings provide population-specific reference data for Turkish adults; however, multicenter studies with independent validation cohorts are warranted to enhance the generalizability of the proposed model.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Approval for this study was obtained from the Ankara University Faculty of Medicine Human Research Ethics Committee (decision number İ04-275-23, dated 10/05/2023), and the principles of the Declaration of Helsinki were taken into consideration.

AUTHORS' CONTRIBUTIONS

EA: Conceptualization, Formal Analysis, Writing – review & editing. **VOS:** Conceptualization, Formal Analysis, Writing – original draft. **MK:** Conceptualization, Data curation, Supervision. **YB:** Conceptualization, Methodology.

DATA AVAILABILITY STATEMENT

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

REFERENCES

1. Dokgöz H. Forensic medicine and forensic sciences [Turkish]. Ankara: Akademisyen Kitabevi; 2019. p. 153.
2. Dokgöz H. Forensic medicine and forensic sciences [Turkish]. Ankara: Akademisyen Kitabevi; 2019. p. 858.
3. Zhang M. The application of forensic imaging to sex estimation: focus on skull and pelvic structures. Perspect Legal Forensic Sc. 2024;1(1):10005. https://doi.org/10.35534/plfs.2024.10005
4. Shim YT, Jeong YH, Aum N, Ha HI, Choi M, Hyun JY, et al. A new metric method for sex estimation using three-dimensional

imaging of the nuchal crest. Anat Cell Biol. 2024;57(4):535-42. https://doi.org/10.5115/acb.24.114

5. Ferrell MJ, Schultz JJ, Adams DM. Sex estimation research trends in forensic anthropology between 2000 and 2022 in five prominent journals. J Forensic Sci. 2024;69(4):1138-54. https://doi.org/10.1111/1556-4029.15522
6. Pilli E, Palamenghi A, Cattaneo C. Forensic skeletal and molecular anthropology face to face: combining expertise for identification of human remains. Ann N Y Acad Sci. 2025;1550(1):77-107. https://doi.org/10.1111/nyas.15398
7. Raghavendra Babu YP, Kanchan T, Attiku Y, Dixit PN, Kotian MS. Sex estimation from foramen magnum dimensions in an Indian population. J Forensic Leg Med. 2012;19(3):162-7. https://doi.org/10.1016/j.jflm.2011.12.019
8. Dereli AK, Zeybek V, Sagtas E, Senol H, Ozgul HA, Acar K. Sex determination with morphological characteristics of the skull by using 3D modeling techniques in computerized tomography. Forensic Sci Med Pathol. 2018;14(4):450-9. https://doi.org/10.1007/s12024-018-0029-0
9. Latham KE, Miller JJ. DNA recovery and analysis from skeletal material in modern forensic contexts. Forensic Sci Res. 2018;4(1):51-9. https://doi.org/10.1080/20961790.2018.1515594
10. Gupta A, Billings BK, Hummel S, Grosskopf B. Evaluating morphological methods for sex estimation on isolated human skeletal materials: comparisons of accuracies between German and South African skeletal collections. Forensic Sci. 2022;2(3):574-84. https://doi.org/10.3390/forensicsci2030042
11. Jakubowska J, Maciejewska A, Pawłowski R. Comparison of three methods of DNA extraction from human bones with different degrees of degradation. Int J Legal Med. 2012;126(1):173-8. https://doi.org/10.1007/s00414-011-0590-5
12. Bhoyar L, Mehar P, Chavali K. An overview of DNA degradation and its implications in forensic caseworks. Egypt J Forensic Sci. 2024;14(15):1-19. https://doi.org/10.1186/s41935-024-00389-y
13. Yasar Teke H, Ünlütürk Ö, Günaydin E, Duran S, Özsoy S. Determining gender by taking measurements from magnetic resonance images of the patella. J Forensic Leg Med. 2018;58:87-92. https://doi.org/10.1016/j.jflm.2018.05.002
14. Gapert R, Black S, Last J. Sex determination from the occipital condyle: discriminant function analysis in an eighteenth and nineteenth century British sample. Am J Phys Anthropol. 2009;138(4):384-94. https://doi.org/10.1002/ajpa.20946
15. Gapert R, Black S, Last J. Sex determination from the foramen magnum: discriminant function analysis in an eighteenth and nineteenth century British sample. Int J Legal Med. 2009;123(1):25-33. https://doi.org/10.1007/s00414-008-0256-0
16. Suazo GIC, Russo PP, Zavando MDA, Smith RL. Sexual dimorphism in the foramen magnum dimensions. Int J Morphol. 2009;27(1):21-3.
17. Hoover KC, Thomas GP. Sexual dimorphism and biomechanical loading in occipital bone morphological variation. Am J Hum Biol. 2022;34(10):e23792. https://doi.org/10.1002/ajhb.23792
18. Ekizoglu O, Hocaoglu E, Inci E, Can IO, Solmaz D, Aksoy S, et al. Assessment of sex in a modern Turkish population using cranial anthropometric parameters. Leg Med (Tokyo). 2016;21:45-52. https://doi.org/10.1016/j.legalmed.2016.06.001
19. Kirisattayakul W, Boonruengudom G, Chalardpongpun P, Duangchit S, Poodendaen C, Aorachon P, et al. Forensic morphometry and prevalence of external occipital protuberance types in Northeastern Thais: CT and dry skull evidence for sex and age estimation. Forensic Sci. 2025;5(4):78. https://doi.org/10.3390/forensicsci5040078
20. Kelley SR, Tallman SD. Population-inclusive assigned-sex-at-birth estimation from skull computed tomography scans. Forensic Sci. 2022;2(2):321-48. https://doi.org/10.3390/forensicsci2020024
21. Meral O, Meydan R, Toklu BB, Kaya A, Karadayi B, Acar T. Estimation of sex from computed tomography images of skull measurements in an adult Turkish population. Acta Radiol. 2022;63(11):1513-21. https://doi.org/10.1177/02841851211044978
22. Zheng B, Zhong Y, Al-Worafi NA, Liu Y. The dimensional and morphological assessment of the frontal sinus in sex estimation among different populations. Head Face Med. 2023;19(1):8. https://doi.org/10.1186/s13005-023-00355-4
23. Luo H, Wang J, Zhang S, Mi C. The application of frontal sinus index and frontal sinus area in sex estimation based on lateral cephalograms among Han nationality adults in Xinjiang. J Forensic Leg Med. 2018;56:1-4. https://doi.org/10.1016/j.jflm.2017.12.014
24. Tsiouris C, Triantafyllou G, Karangeli N, Botis GG, Papadopoulos-Manolarakis P, Kalamatianos T, et al. Morphometric assessment of occipital condyles and foramen magnum reveals enhanced sexual dimorphism detection via 3D imaging: a systematic review and meta-analysis utilizing classification and regression trees. Diagnostics (Basel). 2025;15(11):1359. https://doi.org/10.3390/diagnostics15111359
25. Kizilgoz V, Aydin S, Aydemir H, Keles P, Kantarci M. Interobserver and intraobserver reliability of skull base angles measured on magnetic resonance images. World J Clin Cases. 2024;12(34):6687-95. https://doi.org/10.12998/wjcc.v12.i34.6687

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260628

Systematic analysis of the scientific quality, reliability, and educational value of Vaginal Natural Orifice Transluminal Endoscopic Surgery videos on YouTube

Okay Alptekin[1]*, Betül Kalkan[2]

SUMMARY

OBJECTIVE: The aim of this study was to evaluate the scientific quality, reliability, and educational value of YouTube videos on Vaginal Natural Orifice Transluminal Endoscopic Surgery using structured assessment tools.

METHODS: This cross-sectional study analyzed 20 YouTube videos on Vaginal Natural Orifice Transluminal Endoscopic Surgery identified using the keywords "Vaginal Natural Orifice Transluminal Endoscopic Surgery," "Vaginal Natural Orifice Transluminal Endoscopic Surgery hysterectomy," "vaginal natural orifice endoscopic surgery," and "transvaginal NOTES." Two obstetrics and gynecology specialists independently evaluated the videos using the Global Quality Scale, modified DISCERN, and Journal of the American Medical Association benchmarks. Analyses included non-parametric group comparisons, Spearman correlation, multiple linear regression, and intraclass correlation coefficient analysis.

RESULTS: The median number of views was 2,565 (interquartile range: 13,322). The mean Global Quality Scale, modified DISCERN, and Journal of the American Medical Association scores were 4.17±0.94, 3.95±0.84, and 2.65±0.93, respectively. Inter-rater agreement was moderate for Global Quality Scale and modified DISCERN and good for Journal of the American Medical Association. A statistically significant difference by channel type was observed only for modified DISCERN scores. Although Global Quality Scale scores were moderately correlated with likes (ρ=0.565, p=0.009), multiple linear regression showed no independent association between popularity or technical quality variables and quality scores (p>0.05).

CONCLUSION: Vaginal Natural Orifice Transluminal Endoscopic Surgery videos on YouTube may provide valuable educational content; however, popularity metrics do not reflect scientific reliability. Video selection for surgical education should be based on structured quality assessment tools rather than view count alone.

KEYWORDS: Natural orifice endoscopic surgery. Social media. Video recording. Education, medical. Gynecology. Reproducibility of results.

INTRODUCTION

Vaginal Natural Orifice Transluminal Endoscopic Surgery (vNOTES) provides vaginal access to the peritoneal cavity by combining endoscopic and laparoscopic methods, thereby avoiding an abdominal skin incision and potentially reducing perioperative morbidity[1]. It has been adopted as an alternative to conventional laparoscopy or open surgery in benign gynecologic procedures such as hysterectomy and adnexal surgery[2]. However, limited vaginal access, previous pelvic surgery, and advanced pelvic adhesions may restrict its applicability and require careful patient selection[3].

Surgical education has evolved from the traditional master-apprentice model toward digital learning environments, and YouTube has become one of the most frequently used platforms for health-related and surgical educational content because of its open-access nature[4]. However, the absence of peer review raises concerns regarding accuracy: health-related and surgical videos are frequently incomplete, misleading, or heterogeneous in educational structure, presentation of technical steps, and source transparency[5,6]. Structured scoring systems have therefore been recommended to assess such material reliably[7], and popularity metrics have not been found to reflect scientific quality across surgical fields[8,9].

The literature has nevertheless focused mainly on common procedures, and studies evaluating YouTube content for relatively novel, technically demanding procedures with a pronounced learning curve, such as vNOTES, remain limited. vNOTES has a distinct learning curve due to its different anatomical orientation, limited instrument triangulation, and technical constraints in vaginal access[10].

[1]Erzincan Mengücek Gazi Training and Research Hospital, Department of Obstetrics and Gynecology – Erzincan, Turkey.
[2]Erzincan Binali Yıldırım University, Faculty of Medicine, Department of Obstetrics and Gynecology – Erzincan, Turkey.
*Corresponding author: okayalptekin@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on July 03, 2026. Accepted on August 08, 2026.
Scientific Editor: José Maria Soares Júnior

Studies evaluating the educational value of vNOTES videos remain limited, and the available evidence suggests that such content cannot be considered a reliable standalone resource for surgical education[11]. The present study differs from earlier YouTube analyses in three respects. First, it addresses a procedure still in its diffusion phase, for which structured proctorship is scarce, and trainees therefore rely more heavily on freely available video material. Second, educational quality, scientific reliability, and source transparency were assessed simultaneously with three complementary instruments by two independent and mutually blinded evaluators, with inter-rater agreement quantified rather than assumed. Third, the assumption that popularity indicates quality was tested in a multivariable model rather than by correlation alone. The aim of this study was to evaluate the scientific quality, reliability, and educational value of vNOTES videos on YouTube using structured assessment tools.

METHODS

Video screening strategy

This cross-sectional study was conducted in January 2026 and analyzed YouTube videos using the keywords "vNOTES surgery," "vNOTES hysterectomy," "vaginal natural orifice endoscopic surgery," and "transvaginal NOTES." Searches were performed in incognito mode to minimize bias. The first 50 results sorted by relevance for each keyword were screened. After duplicate removal, the remaining videos were assessed according to inclusion and exclusion criteria. Video selection was performed independently by two researchers, and disagreements were resolved by consensus.

Inclusion and exclusion criteria

The study included videos directly related to vNOTES surgery that contained narration or subtitles in English or Turkish, allowed visualization of surgical steps, had a duration of at least 1 min, more than 100 views, and adequate image quality to clearly identify the surgical field (minimum 720p). Because vNOTES is a relatively recent surgical technique and the amount of available content is limited, a view-count threshold of more than 100 views was applied so that technically relevant videos would not be excluded.

Content intended for advertising or promotion, videos containing only patient experiences or interviews, theoretical presentations without surgical procedures, duplicate videos, and low-quality videos that were silent, lacked subtitles, or did not clearly show the surgical field were excluded. A total of 20 videos were included in the final analysis.

Data collection and assessment scales

For each video, the following data were recorded: duration, upload date, views, likes, comments, language, channel type (academic journal, educational institution, or individual/medical channel), and whether source or reference information was provided in the description section. Data were collected using a standardized form. Video screening and scoring were performed by the two evaluators independently. Each evaluator was blinded to the other's scores, and no scores were shared or discussed until all evaluations were completed and the intraclass correlation coefficients had been calculated.

Both evaluators were blinded to video popularity metrics and scored the videos on content alone. The Global Quality Scale (GQS) assessed overall educational quality on a 1–5 scale, with higher scores indicating better quality.

The modified DISCERN (mDISCERN) scale assessed scientific reliability through five criteria: source citation, impartial presentation, consistency with current scientific evidence, discussion of risks and alternatives, and clarity; each was scored 1 for "yes," giving a total of 0 to 5.

The Journal of the American Medical Association (JAMA) criteria assessed transparency and academic accountability through authorship, source attribution, disclosure of conflicts of interest or sponsorship, and content currency, with one point per criterion and a total of 0 to 4.

Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics version 27 (IBM Corp., Armonk, NY, USA). Continuous variables were assessed using the Shapiro-Wilk test and were not normally distributed ($p<0.05$ for all variables). Therefore, data were presented as mean±standard deviation and median with interquartile range (IQR), and non-parametric tests were used for group comparisons.

The Kruskal-Wallis test was used to compare quality scores by channel type, and significant differences were examined with the Mann-Whitney U test. Relationships between continuous variables were assessed using Spearman's rank correlation coefficient (ρ).

Inter-rater agreement was assessed using the single-measure intraclass correlation coefficient (ICC [2,1]) based on a two-way random-effects model with absolute agreement and reported with 95%CI. ICC values were interpreted as <0.50, poor; 0.50–0.75, moderate; 0.75–0.90, good; and >0.90, excellent.

Multiple linear regression was performed with view count, like count, and video quality as independent variables and GQS, mDISCERN, and JAMA scores as separate dependent variables; multicollinearity, assessed by variance inflation factor, was within acceptable limits.

Two-sided p-values were used in all statistical analyses, and p<0.05 was considered statistically significant.

Ethics statement

Ethics committee approval was obtained from the Erzincan Binali Yıldırım University Non-Interventional Clinical Research Ethics Committee (decision no: 2025-18/06; date: 16.10.2025). The study followed the Declaration of Helsinki and national ethical standards. As it analyzed only publicly available YouTube videos and involved no direct human participation or identifiable personal data, informed consent was not required.

RESULTS

Descriptive statistics and video characteristics

The main characteristics of the 20 included videos are presented in Table 1. The median video duration was 282.5 s (IQR: 290.8; range: 81–1920), the median view count was 2,565 (IQR: 13,322; range: 168–56,650), the median number of likes was 25.5 (IQR: 67.8), and the median number of comments was 0 (IQR: 1.0).

Mean quality scores were 4.17±0.94 for the Global Quality Scale (GQS), 3.95±0.84 for mDISCERN, and 2.65±0.93 for the JAMA criteria (Figure 1).

Seventy-five percent of the videos (n=15) were in English, and 25% (n=5) were in Turkish. According to channel type, educational channels constituted the largest group at 55% (n=11), followed by individual medical channels at 30% (n=6) and academic journal-based channels at 15% (n=3).

Inter-rater reliability (intraclass correlation coefficient)

In the inter-rater reliability analysis between the two independent raters, good reliability was observed for the JAMA scale (ICC=0.706; 95%CI 0.40–0.87; p<0.001). Moderate reliability was observed for the GQS (ICC=0.609; 95%CI 0.25–0.82; p=0.001) and mDISCERN (ICC=0.518; 95%CI 0.12–0.78; p=0.007) scales (Table 2).

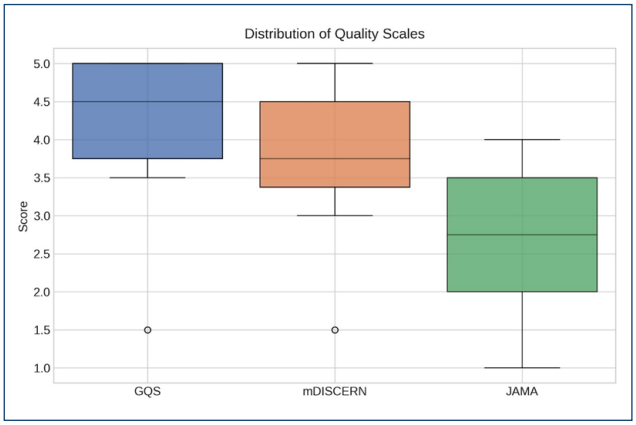


Figure 1. Distribution of Global Quality Scale, modified DISCERN, and Journal of the American Medical Association scores (box plots).

Table 1. General characteristics and descriptive statistics of the videos.

Variable	Mean±SD	Median (IQR)	Min–max
Duration (seconds)	433.2±470.5	282.5 (290.8)	81–1,920
Views	9,095±13,551	2,565 (13,322)	168–56,650
Likes	41.1±45.5	25.5 (67.8)	3–185
Comments	1.1±2.0	0.0 (1.0)	0–6
Mean GQS score	4.17±0.94	4.50 (1.5)	1.5–5.0
Mean mDISCERN score	3.95±0.84	4.00 (1.0)	1.5–5.0
Mean JAMA score	2.65±0.93	2.75 (1.5)	1.0–4.0

SD: standard deviation; IQR: interquartile range; GQS: Global Quality Scale; mDISCERN: modified DISCERN; JAMA: Journal of the American Medical Association.

Table 2. Inter-rater reliability analysis (intraclass correlation coefficient).

Scale	ICC (single absolute)	95%CI	p-value	Level of agreement
GQS	0.609	0.25–0.82	0.001	Moderate
mDISCERN	0.518	0.12–0.78	0.007	Moderate
JAMA	0.706	0.40–0.87	<0.001	Good

ICC: intraclass correlation coefficient; CI: confidence interval; GQS: Global Quality Scale; mDISCERN: modified DISCERN; JAMA: Journal of the American Medical Association.

Comparisons by channel type and language

A statistically significant difference by channel type was observed only for mDISCERN scores (Kruskal-Wallis, p=0.041). Journal-based channels had higher mDISCERN scores than educational and medical channels. However, this group comprised only three videos, and this finding should therefore be interpreted with caution. No significant differences were found for GQS (p=0.083) or JAMA scores (p=0.483). No statistically significant differences were found between Turkish and English videos for GQS, mDISCERN, or JAMA scores (p>0.05 for all comparisons).

Correlation and regression analyses

In the Spearman correlation analysis, a moderate positive correlation was found between GQS scores and views (ρ=0.490, p=0.028) and likes (ρ=0.565, p=0.009). In contrast, no significant relationship was found between mDISCERN and JAMA scores and views, likes, comments, or video quality (p>0.05). In multiple linear regression analysis, none of the models, including view count, like count, and video quality, were significantly associated with GQS, mDISCERN, or JAMA scores. The GQS (R^2=0.312, p=0.104), mDISCERN (R^2=0.238, p=0.214), and JAMA models were all non-significant, and only overall model fit is therefore reported.

DISCUSSION

The findings showed that vNOTES-related YouTube videos generally had good educational quality (GQS: 4.17±0.94) and satisfactory scientific reliability (mDISCERN: 3.95±0.84), but only moderate transparency (JAMA: 2.65±0.93). These findings differ from systematic reviews describing surgical YouTube videos as low quality. Gorgy et al. reported that 96.6% of surgical videos were educationally unsatisfactory[9], and health-related YouTube content in general has been judged to remain below acceptable standards[4]. In the only previous vNOTES-specific analysis, Temur reported that 43.1% of videos were low quality and that the median GQS score was 2[11].

This may be explained by the inclusion criteria applied in our study, which required a minimum resolution of 720p, a minimum duration of 1 min, and clear visualization of the surgical field. These criteria may have improved the overall quality of the sample by excluding low-resolution or amateur content that could contribute to misinformation. In addition, 15% of the videos were sourced from academic journals and 55% from structured educational channels, which may have contributed to the higher reliability observed in journal-based channels (p=0.041). However, only three videos originated from journal-based channels, and this observation should therefore be regarded as preliminary. Comparable observations have been reported for other gynecologic procedures, where videos uploaded predominantly by healthcare professionals nevertheless achieved only moderate overall quality and low mDISCERN scores because of insufficient academic referencing[12,13]. These findings suggest that high visual quality, reflected by GQS, does not always correspond to high scientific reliability, reflected by mDISCERN.

Videos produced by academic institutions, professional organizations, or surgeons have consistently been reported to provide higher educational value and transparency than those from individual users[6,9,11].

One of the main findings of our study was that view and like counts had no independent effect on quality scores in multiple linear regression analysis. Assuming that the most viewed video is also the most reliable may therefore be misleading, as platform algorithms prioritize engagement over accuracy and allow scientifically flawed techniques to reach large audiences[4]. Therefore, structured assessment tools such as GQS, mDISCERN, and JAMA should be prioritized over popularity metrics in surgical educational materials.

vNOTES has a distinct learning curve because of its reverse anatomical perspective and limited instrument movement, with approximately 16 cases required for operative stability and 18 for proficiency in vNOTES hysterectomy[10]. However, Kapurubandara et al. emphasized that attempting innovative techniques such as vNOTES based solely on YouTube videos may involve serious surgical risks without structured proctorship training[3]. The educational value of video materials depends on clear demonstration of surgical safety, including the Critical View of Safety; however, compliance with these standards has generally been reported to be below 10%[9].

The study was limited by the small sample of 20 videos, which restricts generalizability, and by the dynamic nature of YouTube content, meaning that the data reflect a single time point[11]. The sample was small because the number of vNOTES videos meeting the predefined technical and content criteria was limited; findings based on 20 videos, and in particular subgroup comparisons involving as few as three videos, should therefore be regarded as descriptive and hypothesis-generating rather than definitive. In addition, the inclusion criteria applied here (minimum 720p resolution, minimum 1-min duration, and clear visualization of surgical steps) are likely to have preferentially selected structured, institutional, or professionally produced content. The analyzed sample therefore represents a preselected, comparatively high-quality subset rather than the full spectrum of vNOTES content on YouTube. This may explain the higher

scores observed compared with broader YouTube analyses and limits the generalizability of our findings. Peer-reviewed official video channels created by surgical societies and academic institutions may improve digital medical education and help counter anecdotal bias on YouTube[4]. For surgical education, these findings support a practical approach: video material should be selected by structured criteria rather than view count, preferably from accredited society or journal channels; it should complement rather than replace proctored hands-on training; and trainees should verify authorship, referencing, and disclosure before adopting a technique observed online. Curated, peer-reviewed vNOTES video libraries in which safety-critical steps are explicitly demonstrated would support this.

CONCLUSION

Within this preselected sample, YouTube videos on vNOTES provided visually adequate educational content, although quality varied considerably. Videos from academic journal-based sources tended to show the highest reliability scores; however, this group comprised only three videos and the observation requires confirmation in larger samples. Because view and like counts are insufficient indicators of quality, structured assessment tools such as GQS and DISCERN should be used in selecting videos for surgical education. As our inclusion criteria yielded a technically preselected subset, these findings cannot be generalized to all vNOTES content on YouTube. Accredited channels endorsed by surgical societies may further improve the reliability of digital medical education.

AUTHORS' CONTRIBUTIONS

OA: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **BK:** Funding acquisition, Investigation, Resources, Supervision, Validation, 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. Wylie S, Mutema E, Yoong W. Vaginal natural orifice transluminal endoscopic surgery (vNOTES): a new chapter in vaginal surgical innovation. J Obstet Gynaecol. 2025;45(1):2458842. https://doi.org/10.1080/01443615.2025.2458842
2. Marchand GJ, Ulibarri H, Arroyo A, Blanco M, Herrera DG, Hamilton B, et al. Systematic review and meta-analysis of vaginal natural orifice transluminal endoscopic surgery hysterectomy versus vaginal hysterectomy for benign indications. AJOG Glob Rep. 2024;4(2):100355. https://doi.org/10.1016/j.xagr.2024.100355
3. Kapurubandara S, Lowenstein L, Salvay H, Herijgers A, King J, Baekelandt J. Consensus on safe implementation of vaginal natural orifice transluminal endoscopic surgery (vNOTES). Eur J Obstet Gynecol Reprod Biol. 2021;263:216-22. https://doi.org/10.1016/j.ejogrb.2021.06.019
4. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. BMC Med Educ. 2022;22(1):382. https://doi.org/10.1186/s12909-022-03446-z
5. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J. 2015;21(3):173-94. https://doi.org/10.1177/1460458213512220
6. Noll D, Green L, Asokan G, Barbaro A, Hewitt J, Ovenden C, et al. Is YouTube™ a good source of information for patients to understand laparoscopic fundoplication? Ann R Coll Surg Engl. 2023;105(4):365-71.
7. de'Angelis N, Gavriilidis P, Martínez-Pérez A, Genova P, Notarnicola M, Reitano E, et al. Educational value of surgical videos on YouTube: quality assessment of laparoscopic appendectomy videos by senior surgeons vs. novice trainees. World J Emerg Surg. 2019;14:22. https://doi.org/10.1186/s13017-019-0241-6
8. Erdem MN, Karaca S. Evaluating the accuracy and quality of the ınformation in kyphosis videos shared on YouTube. Spine (Phila Pa 1976). 2018;43(22):E1334-9. https://doi.org/10.1097/brs.0000000000002691
9. Gorgy A, Hawary HE, Galli R, MacDonald M, Barone N, Thibaudeau S. Evaluating the educational quality of surgical YouTube® videos: a systematic review. Health Sci Rev. 2022;5:100067. https://doi.org/10.1016/j.hsr.2022.100067
10. Housmans S, Baekelandt J, Deprest J. Proctored single surgeon learning curve for vaginal natural orifice transluminal endoscopic surgery (NOTES) hysterectomy. Eur J Obstet Gynecol Reprod Biol. 2025;307:223-9. https://doi.org/10.1016/j.ejogrb.2025.02.020
11. Temur I. Evaluation of the educational value, reliability, and quality of vaginally-assisted natural orifice transluminal endoscopic surgery (v-NOTES) videos on YouTube. Cureus. 2025;17(6):e85244. https://doi.org/10.7759/cureus.85244
12. Erdogan G. Female genital cosmetic surgery (FGCS): evaluation of YouTube videos. J Gynecol Obstet Hum Reprod. 2021;50(4):102102. https://doi.org/10.1016/j.jogoh.2021.102102
13. Özcan C, Sancı A. Evaluation of robotic-assisted sacrocolpopexy videos on YouTube. Int Urogynecol J. 2023;34(8):1843-7. https://doi.org/10.1007/s00192-023-05480-x

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260634

Evaluation of the ranawat triangle method for determining hip rotation center in total hip arthroplasty among Turkish patients

Alper Dünki[1]*, Hakan Batuhan Kaya[1], Murat Birinci[1], Furkan Uçar[1], Ömer Polat[1], Savaş Çamur[1]

SUMMARY

OBJECTIVE: Accurate identification of the anatomical femoral head center is critical in total hip arthroplasty for optimal implant positioning and long-term survival. The Ranawat triangle method is widely used for this purpose, but its applicability may vary across populations. The aim of this study was to evaluate the accuracy and suitability of the Ranawat triangle method in estimating the anatomical femoral head center in a Turkish population.

METHODS: Standardized anteroposterior pelvic radiographs of 260 non-fractured hips from 312 patients aged 60–90 were retrospectively analyzed. Pelvic height, pelvic width, and femoral head diameter were measured. The coordinates of the anatomical femoral head center were calculated anatomically and using the Ranawat triangle method. Deviations between the two were assessed in both axes, with <5 mm considered acceptable error.

RESULTS: Anatomical femoral head center estimated by the Ranawat triangle method was positioned an average of 4.7±1.5 mm medially and 1.0±0.8 mm cranially in men, and 4.8±1.1 mm medially and 1.1±0.7 mm cranially in women, relative to the anatomical center. While 100% of Y-axis errors and 58.1% of X-axis errors were within 5 mm. Strong positive correlations were found between pelvic height and Cy (r=0.59), and pelvic width and Cx (r=0.82).

CONCLUSION: The Ranawat triangle method method demonstrated excellent accuracy in predicting the vertical (Y-axis) position of the anatomical femoral head center but showed moderate lateral (X-axis) deviations in over 40% of cases. Despite its limitations in lateral offset prediction, it may serve as a practical reference method for preoperative planning, particularly for estimating the vertical position of the hip rotation center in Turkish total hip arthroplasty patients.

KEYWORDS: Arthroplasty. Hip joint. Anatomic landmarks. Radiometry.

INTRODUCTION

Increasing life expectancy and the rising prevalence of osteoarthritis have led to a marked growth in the number of total hip arthroplasty procedures performed. Although this technique is highly successful, it is not entirely free of complications. Periprosthetic joint dislocations are the most common mechanical complications and are usually related to component placement[1,2]. Acetabular and femoral component position, combined anteversion, acetabular inclination, and femoral stem varus being appropriate for anatomic placement are important for long-term survival[3,4].

The femoral head rotation center may not always be easily determined due to degeneration of the hip joint. For this reason, various methods have been determined and the Ranawat triangle method (RTM) is one of the most frequently used ones[5].

Pelvic length, width, and anatomical femoral head center (AFHC) may vary according to race and gender[6]. Accordingly, there is no publication in the literature showing the applicability of the Turkish population with the RTM in determining AFHC. The aim of the study is to determine how accurately the RTM estimates in determining AFHC in Turkish patients.

METHODS

Following Institutional Review Board (IRB) approval, pelvic radiographs of 260 patients were retrospectively reviewed. Radiological assessments were performed utilizing standardized anteroposterior (AP) pelvic radiographs. To establish a consistent reference system, orthogonal X- and Y-axes were defined at the outset. The X-axis was designated as the mediolateral line parallel to the plane connecting the most inferior margins of the bilateral teardrop figures, whereas the Y-axis was determined as the craniocaudal line drawn perpendicular to the X-axis, thereby forming a Cartesian coordinate system[6].

In hips without radiographic evidence of arthrosis, the anatomical hip rotation center was identified through geometric

[1]Umraniye Training and Research Hospital, Department of Orthopaedics and Traumatology – Istanbul, Turkey.
*Corresponding author: alperdunki@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on June 25, 2026. Accepted on August 15, 2026.
Scientific Editor: José Maria Soares Júnior

fitting. Specifically, a best-fit circle was drawn to approximate the femoral head contour, and the circle's center was designated as the anatomical rotation point. This center was subsequently used as the primary reference for all coordinate-based calculations, expressed as Cx and Cy along the X- and Y-axes, respectively[7].

Pelvic width (PW) was measured on the X-axis as the distance between the lateral edges of the iliac crests. Pelvic height (PH) was defined on the Y-axis, extending from the highest point of the iliac crest to the lowest point of the ischial tuberosity. The vertical femoral head diameter (FHD) was also obtained on the Y-axis as the distance from the femoral head apex to its junction with the neck.

The AFHC was estimated using the method originally described by Ranawat et al.[8]. This technique approximates the true acetabular region by constructing an isosceles right triangle, in which both the base and height correspond to 20% of the measured PH. The triangle was positioned such that its inferior margin originated at the intersection of Kohler's line and Shenton's line. The AFHC was defined as the midpoint of the hypotenuse of this constructed triangle, and its coordinates were designated as Ax and Ay (Figure 1).

To evaluate anatomical alignment and method accuracy, several ratios and differences were calculated, including FHD/PH, Cx/PH, Cy/PH, Ax–Cx, and Ay–Cy. The RTM was considered reliable when the absolute differences between estimated and anatomical centers (Ax–Cx and Ay–Cy) were within a 5 mm threshold[6].

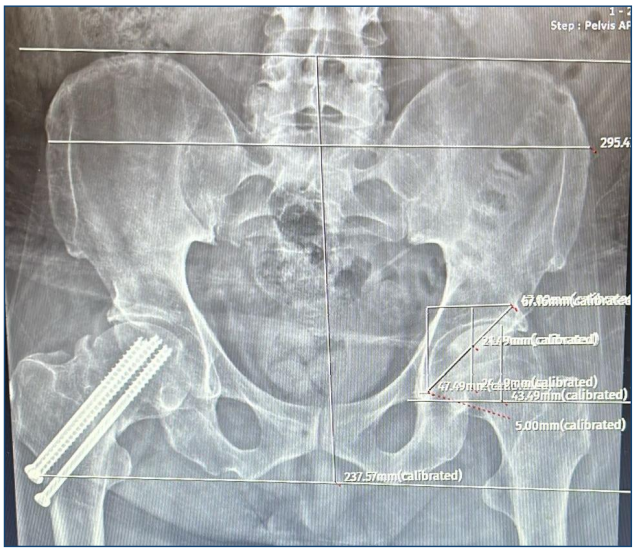


Figure 1. Measurement of the hip center of rotation on a calibrated anteroposterior pelvic radiograph using the Ranawat triangle method.

All radiographic measurements were repeated twice by the first author with at least a 2-week interval to minimize intraobserver variation. Intraobserver reliability of all radiographic measurements was evaluated using the intraclass correlation coefficient (ICC). The analysis demonstrated an ICC value of 0.88, reflecting excellent reliability of the measurement technique. Data normality and variance homogeneity were tested with Shapiro-Wilk and F-tests. Student's t-test or Wilcoxon rank-sum was applied for continuous variables, while categorical variables were assessed using the chi-square test. Associations were evaluated with Pearson correlation. Statistical significance was set at $p<0.05$, and analyses were performed with Statistical Package for the Social Sciences (SPSS) 29.0 (IBM Corp., Armonk, NY, USA).

Level of evidence: Level IV—Retrospective Observational Study.

RESULTS

The mean Cx and Cy values were found to be 23.5±2.2 mm and 23.0±2.1 mm, respectively. Mean Cx, Cy, FHD, and PW were significantly greater in males compared to females, whereas no significant difference was observed in PH between sexes (Table 1). While the Cx/PH and Cy/PH ratios showed no significant gender-based differences, the FHD/PH ratio was higher in males than in females.

The statistical evaluation indicated that PH showed a significant positive association with both coordinate values and femoral head morphology. Specifically, PH was positively related to the horizontal component (Cx, $r=0.58$, $p<0.001$), the vertical component (Cy, $r=0.59$, $p<0.001$), and the vertical FHD ($r=0.58$, $p<0.001$). PW also demonstrated strong associations with the same parameters. Higher PW values were linked to greater Cx ($r=0.82$, $p<0.001$), Cy ($r=0.80$, $p<0.001$), and FHD ($r=0.72$, $p<0.001$) measurements.

Application of the RTM yielded an estimated AFHC with mean coordinates of 28.3±1.8 mm along the Ax and 23.3±1.8 mm along the Ay. The mean deviation from the anatomical reference point was 4.7±1.3 mm in the mediolateral direction and 1.0±0.8 mm in the craniocaudal direction. These results demonstrate that the AFHC was consistently located medial and cranial to the true anatomical hip rotation center.

A subgroup analysis revealed sex-specific differences. In males, the AFHC was located approximately 4.7±1.5 mm medial and 1.0±0.8 mm cranial to the anatomical reference. In females, the displacement was similar, averaging 4.8±1.1 mm medially and 1.1±0.7 mm cranially relative to the true hip center (Table 1).

Regarding estimation performance, the mediolateral measurements exhibited substantial variability, with only 5.4% of

all evaluated hips falling within a 3-mm range. In contrast, the craniocaudal assessments demonstrated markedly higher precision, as 98.8% of hips were within the same margin. When the threshold was extended to 5 mm, a distinct disparity between the two axes became apparent: 58.1% of hips fell within this limit along the X-axis, whereas all hips remained below the 5-mm deviation on the Y-axis (Table 2).

DISCUSSION

Preoperative templating of the hip joint rotation center is important for long-term survival and reducing complications in hip arthroplasty. Various methods have been described to determine the correct hip rotation center in pelvic AP radiographs. However, due to morphological differences in population, not every method is suitable for every society[9]. In this study, the applicability of the RTM was examined in Turkish individuals with normal hip morphology. Measurements along the Y-axis remained within 5 mm in all patients. In contrast, 41.9% of hips demonstrated a displacement greater than 5 mm on the X-axis.

Takamatsu et al. found that the hip rotation center aligned with PH on the Y-axis and with PW on the X-axis[10]. In a similar study conducted with 100 THA patients in 2019, according to Yi et al., the position of the hip rotation center along the X- and Y-axes showed a positive association with PW and PH in Chinese individuals[11]. In a study conducted by Bombaci et al. in 2017 with 100 patients aged between 18 and 91, a positive correlation was observed between the distance of the location of the hip rotation center on the X-axis to the ischial tuberosity and the ratio of PW; the distance of the location on the Y-axis to the teardrop figure and PH in Turkish population[12].

Fujii et al. reported that Cx showed a weak correlation with PW in the Japanese population in their study with 123 patients, but there was a strong correlation between Cy and PH. Fujii et al., who also determined the hip rotation center

Table 1. Demographic and radiographic parameters.

	Men (n: 96)	Women (n: 164)	Statistical test	p
Age (y)	76.2±7.8	77.0±7.7	Mann-Whitney U	0.416
Laterality (L/R)	46/50	73/91	Chi-square test	0.595
LCEA (°)	46.1±9.2	45.6±10.5	Student's t-test	0.698
Cx (mm)	24.3±2.4	23.1±2.0	Mann-Whitney U	<0.001
Cy (mm)	23.8±2.0	22.5±2.1	Student's t-test	<0.001
PH (cm)	33.5±2.5	33.8±2.9	Student's t-test	0.372
PW (cm)	24.0±1.9	22.9±1.6	Student's t-test	<0.001
FHD (mm)	49.2±4.6	45.5±3.9	Student's t-test	<0.001
Cx/PH	10.1±0.6	10.0±0.4	Mann-Whitney U	0.398
Cy/PH	9.9±0.5	9.8±0.5	Student's t-test	0.318
FHD/PH	20.5±1.4	19.8±1.2	Mann-Whitney U	<0.001

cm: centimeter; FHD: femoral head diameter; mm: millimeter; n: sample size; PH: pelvic height; PW: pelvic width; y: year; °: angle; LCEA: lateral center edge angle.

Table 2. Morphologic parameters.

	Men (n: 96)	Women (n: 164)	Statistical test	p
Ax (mm)	29.0±1.9	27.9±1.6	Student's t-test	<0.001
Ax–Cx (mm)	4.7±1.5	4.8±1.1	Mann-Whitney U	0.401
|Error| <3 mm	8 hips	6 hips	Chi-square test	0.107
|Error| <5 mm	59 hips	92 hips	Chi-square test	0.398
Ay (mm)	24.0±1.9	22.9±1.6	Student's t-test	<0.001
Ay–Cy (mm)	1.0±0.8	1.1±0.9	Mann-Whitney U	0.422
|Error| <3 mm	93 hips	160 hips	Chi-square test	0.742
|Error| <5 mm	96 hips	164 hips	–	–

mm: millimeter; n: sample size.

according to the RTM, found the estimation error in the X-axis as 6.7±2.6 mm and in the Y-axis as 6.6±1.9 mm. The accuracy of the estimations demonstrated that the absolute error values remained within a tolerance margin of 5 mm along both coordinate axes in a substantial proportion of cases, specifically in 118 hips (96%) for the X-axis and in 116 hips (94%) for the Y-axis, respectively[6]. The current study evaluates the hip rotation center in the Turkish population, and a strong positive correlation of Cx with PW and Cy with PH was observed. According to the RTM, 58.1% of the hips in the X-axis and Y-axis were below 5 mm in all patients, and the obtained data are consistent with the literature.

In examining the relationship between FHD and PH, Crowe et al. identified the FHD/PH ratio to be approximately 20%, forming the basis for their classification of hip dysplasia[13]. Similarly, Zhang et al, in a 2014 study involving 187 Chinese patients, determined the mean FHD/PH ratio to be 21.5% and concluded that the Chinese population also conforms to the Crowe classification[14]. In the present study, the FHD/PH ratio was measured as 20.5±1.4% in men and 19.8±1.2% in women, indicating that the Turkish population likewise aligns with the Crowe classification.

Various patient-specific methods have been described in the literature, including the RTM[8,15-17]. Schofer et al. evaluated the teardrop figure with the current method they applied to 115 patients in 2010. They stated that there was a non-negligible deviation with the RTM[18]. John et al. described in their study that the hip rotation center shifted medially and proximally with the RTM[19]. Fujii et al. also observed medial and proximal shifts in their study and described the modified RTM. In the current study, it was observed that the RTM had a negligible error in the Y-axis (mean 1.0±0.8 mm) and we observed that the hip rotation center shifted medially in the X-axis (mean 4.7±1.2 mm).

The study had some limitations. First, the study was conducted on patients who were candidates for THA in the Turkish population between the ages of 60 and 90. Second, the height, weight, and body mass index of the patients are unknown. Accordingly, PH, PW, and FHD values may vary. Finally, no matter how standardized the radiographs are, pelvic tilt and sacral slope have different values in each patient and AFHC coordinates, PH and PW values may vary in standard pelvic AP radiographs.

CONCLUSION

The hip rotation center remained below 5 mm deviation in the Y-axis in all patients. Thus, the RTM showed considerable accuracy in predicting AFHC in THA cases. However, it showed poor accuracy in determining the lateral offset, with 58.1% deviation below 5 mm in the X-axis.

The RTM may be useful, albeit limited, in determining the hip rotation center in the placement of the acetabular component in THA cases applied to Turkish patients. Therefore, while the RTM can be reliably used to estimate the vertical position of the AFHC, surgeons should be aware of its tendency to underestimate the lateral offset and consider intraoperative or alternative radiographic adjustments when planning acetabular component placement. Given the frequent underestimation of lateral offset, precise preoperative template adjustments are essential to prevent postoperative gait dysfunction and joint instability. Clinicians should thus combine the RTM with anatomical landmarks to accurately restore proper lateral positioning.

ACKNOWLEGMENTS

Ethical approval for the study was obtained from the Ümraniye Training and Research Hospital Ethics Committee on 12 March 2025 (Ref No: E-54132726-000-284739358).

AUTHORS' CONTRIBUTIONS

AD: Conceptualization, Data curation, Writing – review & editing. **HBK:** Software, Validation. **MB:** Data curation. **FU:** Data curation. **ÖP:** Methodology, Writing – review & editing. **SÇ:** 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. Paterno SA, Lachiewicz PF, Kelley SS. The influence of patient-related factors and the position of the acetabular component on the rate of dislocation after total hip replacement. J Bone Joint Surg Am. 1997;79(8):1202-10. https://doi.org/10.2106/00004623-199708000-00013
2. Dunki A, Birinci M, Kaya HB, Akkaya M, Polat O. The role of arthroplasty surgery experience on successful outcomes in primary total hip arthroplasty surgery. North Clin Istanb. 2026;13(2):194-201. https://doi.org/10.14744/nci.2025.77854
3. Radaelli M, Buchalter DB, Mont MA, Schwarzkopf R, Hepinstall MS. A new classification system for cementless femoral stems

in total hip arthroplasty. J Arthroplasty. 2023;38(3):502-10. https://doi.org/10.1016/j.arth.2022.09.014

4. Alarcon Perico D, Warne CN, Lee SH, Roberts HJ, Sierra RJ. Combined anteversion technique for total hip arthroplasty with handheld accelerometer-based navigation system. Arthroplast Today. 2023;23:101193. https://doi.org/10.1016/j.artd.2023.101193
5. Schofer I, Pressel T, Heyse TJ, Schmitt J, Boudriot U. Radiological determination of the anatomic hip centre from pelvic radiographs: comparison of methods. Acta Orthop Belg. 2010;76(4):425-32.
6. Fujii M, Nakamura T, Hara T, Nakashima Y. Is Ranawat triangle method accurate in estimating hip joint center in Japanese population? J Orthop Sci. 2021;26(2):219-24. https://doi.org/10.1016/j.jos.2020.03.007
7. Dandachli W, Nakhla A, Iranpour F, Kannan V, Cobb JP. Can the acetabular position be derived from a pelvic frame of reference? Clin Orthop Relat Res. 2009;467(4):886-93.
8. Ranawat CS, Dorr LD, Inglis AE. Total hip arthroplasty in protrusio acetabuli of rheumatoid arthritis. J Bone Joint Surg Am. 1980;62(7):1059-65.
9. Zacharia B, Fawas KM. A comparative radiographic morphometric analysis to assess the normal radiological morphology of the adult hip in indian population. J Clin Orthop Trauma. 2020;15:117-24. https://doi.org/10.1016/j.jcot.2020.10.015
10. Takamatsu T, Shishido T, Takahashi Y, Masaoka T, Tateiwa T, Kubo K, et al. Radiographic determination of hip rotation center and femoral offset in Japanese adults: a preliminary investigation toward the preoperative implications in total hip arthroplasty. Biomed Res Int. 2015;2015:610763. https://doi.org/10.1155/2015/610763
11. Yi LH, Li R, Zhu ZY, Bai CW, Tang JL, Zhao FC, et al. Anatomical study based on 3D-CT image reconstruction of the hip rotation center and femoral offset in a Chinese population: preoperative implications in total hip arthroplasty. Surg Radiol Anat. 2019;41(1):117-24. https://doi.org/10.1007/s00276-018-2143-9
12. Bombaci H, Simsek B, Soyarslan M, Murat Yildirim M. Determination of the hip rotation centre from landmarks in pelvic radiograph. Acta Orthop Traumatol Turc. 2017;51(6):470-3. https://doi.org/10.1016/j.aott.2017.09.004
13. Crowe JF, Mani VJ, Ranawat CS. Total hip replacement in congenital dislocation and dysplasia of the hip. J Bone Joint Surg Am. 1979;61(1):15-23.
14. Zhang Y, Jiang J, Wang C, Zuo J, Yang M, Wang J. The ratio of femoral head diameter to pelvic height in the normal hips of a Chinese population. Eur J Orthop Surg Traumatol. 2014;24(6):947-51.
15. Fessy MH, N'Diaye A, Carret JP, Fischer LP. Locating the center of rotation of the hip. Surg Radiol Anat. 1999;21(4):247-50.
16. Pierchon F, Migaud H, Duquennoy A, Fontaine C. Radiologic evaluation of the rotation center of the hip. Rev Chir Orthop Reparatrice Appar Mot. 1993;79(4):281-4.
17. Boudriot U, Hilgert J, Hinrichs F. Determination of the rotational center of the hip. Arch Orthop Trauma Surg. 2006;126(6):417-20.
18. Schofer MD, Pressel T, Heyse TJ, Schmitt J, Boudriot U. Radiological determination of the anatomic hip centre from pelvic landmarks. Acta Orthop Belg. 2010;76(4):479-85.
19. John JF, Fisher PE. Radiographic determination of the anatomic hip joint center. Acta Orthop Scand. 1994;65(5):509-10.

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260659

Gastrin-releasing peptide, neuromedin B, and symptom severity in adult attention-deficit hyperactivity disorder

Elif Özcan Tozoğlu[1]*, Alev Lazoğlu Özkaya[2]

SUMMARY

OBJECTIVE: Attention-deficit hyperactivity disorder is a neurodevelopmental disorder associated with metabolic and neuropeptidergic processes. Gastrin-releasing peptide and neuromedin B, which are bombesin-like peptides, have been studied only to a limited extent in adult attention-deficit hyperactivity disorder, despite their role in energy homeostasis and central nervous system function. The aim of this study was to evaluate gastrin-releasing peptide and neuromedin B levels in untreated adult attention-deficit hyperactivity disorder patients and to examine their relationship with clinical symptom severity.

METHODS: This cross-sectional study included 55 untreated adult attention-deficit hyperactivity disorder patients and 32 healthy controls. Attention-deficit hyperactivity disorder diagnoses were confirmed using SCID-5/CV and DIVA 2.0, and symptom severity was assessed using the Adult Attention-Deficit Hyperactivity Disorder Self-Report Scale. Gastrin-releasing peptide and neuromedin B levels in fasting serum were measured using the ELISA method. Group comparisons, Pearson correlations, hierarchical regression, and receiver operating characteristic analyses were performed.

RESULTS: Gastrin-releasing peptide (22.83±24.18 pg/mL) and neuromedin B (1.83±2.86 pg/mL) levels were significantly lower in the attention-deficit hyperactivity disorder group compared to the control group ($p=0.004$ and $p=0.037$, respectively). Gastrin-releasing peptide and neuromedin B levels were negatively correlated with Attention-Deficit Hyperactivity Disorder Self-Report Scale total and subscale scores ($p<0.001$ for all). In regression analyses, gastrin-releasing peptide was an independent predictor of Attention-Deficit Hyperactivity Disorder Self-Report Scale total score ($\beta=-0.642$, $p=0.024$). In receiver operating characteristic analyses, gastrin-releasing peptide (AUC=0.746) and neuromedin B (AUC=0.719) showed significant performance in distinguishing attention-deficit hyperactivity disorder.

CONCLUSION: The findings suggest that gastrin-releasing peptide, in particular, may be associated with clinical symptom severity in adult attention-deficit hyperactivity disorder and is a potential biomarker candidate.

KEYWORDS: Attention deficit disorder with hyperactivity. Biomarkers. Gastrin-releasing peptide. Neuromedin b. Neuropeptides.

INTRODUCTION

Attention-deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder characterized by inattention, hyperactivity, and impulsivity, which may persist into adulthood. Its etiology is thought to involve complex interactions between genetic, environmental, and neurobiological factors[1].

Recent research has focused on the role of metabolic processes and energy homeostasis in neurodevelopmental disorders, suggesting that appetite-related neuropeptides may contribute to behavioral and neurobiological mechanisms associated with ADHD.

Bombesin-like peptides, including gastrin-releasing peptide (GRP) and neuromedin beta (NMB), are anorexigenic neuropeptides that regulate metabolic and neurobehavioral processes via specific G-protein-coupled receptors[2,3]. Experimental studies suggest that GRP and NMB are involved in metabolic regulation and energy balance[4-6].

In addition to their metabolic functions, GRP and NMB are involved in central nervous system processes related to stress responses, synaptic plasticity, and behavioral regulations[7,8]. Clinical studies in autism have reported improvements in symptoms such as hyperactivity and social functioning following GRP administration[9,10]. Moreover, experimental findings indicate that GRP can stimulate dopamine and norepinephrine release, both of which are neurotransmitter systems associated with ADHD[11].

These neuropeptides are also associated with metabolic and inflammatory processes, including glucose metabolism, lipid regulation, and immune responses[12-14]. Therefore, evaluating GRP and NMB levels alongside metabolic parameters

[1]University of Health Sciences, Erzurum Faculty of Medicine, Department of Psychiatry – Erzurum, Türkiye.
[2]Erzurum City Hospital, Department of Biochemistry – Erzurum, Turkey.
*Corresponding author: drelifozcan@hotmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on June 03, 2026. Accepted on August 04, 2026.
Scientific Editor: José Maria Soares Júnior

may contribute to a more comprehensive understanding of homeostatic mechanisms.

Despite growing interest in neuropeptide-based mechanisms, there are no clinical studies evaluating GRP and NMB levels together in untreated adult ADHD populations. Therefore, the present study aimed to investigate GRP and NMB levels in adults with ADHD and examine their relationship with symptom severity while controlling for potential metabolic and inflammatory confounders.

METHODS

Ethics approval and consent to participate

Ethical approval for this study was obtained from the Erzurum Faculty of Medicine Scientific Research Ethics Committee, University of Health Sciences (Decision No: 2024/12-216). The study was conducted in accordance with the Declaration of Helsinki, and written informed consent was obtained from all participants.

Research sample

The sample consisted of untreated adult ADHD patients who presented to the Psychiatry Outpatient Clinic of Erzurum City Hospital between December 2024 and May 2025, and age- and gender-matched healthy controls without psychiatric disorders.

Sample size was calculated using G*Power 3.1 (α=0.05, power=0.80, effect size d=0.5), indicating a minimum of 64 participants.

Participant selection and exclusion criteria

The study included adults aged 18–65 diagnosed with ADHD who had not received prior psychiatric treatment, and healthy controls matched for age and gender. ADHD diagnosis was confirmed using SCID-5/CV and DIVA 2.0, and individuals with additional psychiatric disorders were excluded.

To minimize confounding metabolic and inflammatory effects on GRP and NMB levels, the following exclusion criteria were applied: BMI ≥30 kg/m^2, history of diabetes, hypertension, liver or renal disease, active infection, use of lipid-lowering medications, and severe dyslipidemia (LDL>190 mg/dL or triglycerides>400 mg/dL). Mild-to-moderate lipid elevations below these thresholds were allowed and controlled for as covariates.

Procedure

This cross-sectional study evaluated clinical characteristics, biochemical parameters, and GRP and NMB levels. Symptom severity was assessed using the Adult ADHD Self-Report Scale (ASRS), and sociodemographic data were collected via a standardized form.

Venous blood samples were collected between 8:00 and 10:00 a.m. after an 8–12-h overnight fast. Participants were instructed to avoid alcohol, caffeine, and intense physical activity prior to sampling. Blood samples from female participants were collected between days 2 and 5 of the menstrual cycle to minimize hormonal variability.

Data collection tools

Sociodemographic data were collected using a structured form. Psychiatric diagnoses were assessed using SCID-5/CV, and ADHD diagnosis was supported by DIVA 2.0[15]. Symptom severity was evaluated using the ASRS[16,17].

Biochemical analysis

Serum samples were obtained following centrifugation and stored at -80°C until analysis. Routine biochemical parameters (ALT, AST, GGT, glucose, lipid profile, HbA1c) were measured using standard automated laboratory methods.

GRP and NMB levels were determined using commercially available ELISA kits according to manufacturer protocols.

Statistical analysis

Data were analyzed using SPSS 26.0. Continuous variables were expressed as mean±SD, and categorical variables as frequencies and percentages. Group comparisons were performed using independent samples t-tests and chi-square tests.

Pearson correlation analysis was used to assess relationships between biomarkers and symptom severity. Hierarchical multiple regression analyses were conducted to evaluate predictors of ADHD symptoms, with age, gender, smoking, LDL, HbA1c, and C-reactive protein (CRP) entered in the first block and GRP and NMB in the second block.

Receiver operating characteristic (ROC) analysis was used to evaluate the diagnostic performance of GRP and NMB; area under the curve (AUC), optimal cutoff values (Youden index), sensitivity, and specificity were calculated. A p-value<0.05 was considered statistically significant.

RESULTS

A total of 87 participants were included in the study, comprising 55 ADHD patients and 32 healthy controls. This sample size was found to meet the minimum requirement determined by G*Power. The sociodemographic data of the participants are presented in Table 1. There was no significant difference between the ADHD and control groups in terms of age and gender (p>0.05).

When the biochemical measurements of the groups were evaluated, GRP and NMB levels were found to be significantly lower in the ADHD group compared to the control group (p=0.004 and p=0.037, respectively; Table 1). Pearson correlation analysis revealed a strong positive correlation between GRP and NMB levels (r=0.914, p<0.001). Both biomarkers were significantly and negatively correlated with ASRS total and subscale scores (all p<0.001), indicating that lower biomarker levels were associated with greater symptom severity.

In hierarchical multiple regression analyses conducted to evaluate the predictors of ADHD symptoms (Table 2), the model created for the ASRS total score, with age, gender, smoking, LDL, HbA1c, and CRP in the first block, explained 35.9% of the variance. Adding GRP and NMB in the second block significantly improved the model and increased the total variance explained to 52.5%. In this model, only GRP was found to be a significant predictor (β=-0.642, p=0.024), while the independent effect of NMB was not significant.

The analysis for the attention-deficit subscale showed that the first block explained 34.2% of the variance. Adding GRP and NMB in the second block significantly increased the variance explained, and the total variance explained reached 53.7%.

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

Variable		ADHD (n, mean±SD)	Control (n, mean±SD)	p-value/χ^2
Age		32.01	32.96	0.718
Gender	Female	25	14	0.877
	Male	30	18	
Education	Primary education	17	0	0.001
	High school	15	8	
	University	23	24	
Marital status	Married	28	22	0.267
	Single	24	9	
	Divorced/widowed	3	1	
Occupation/employment status	Worker-employee	26	26	0.008
	Student	4	2	
	Housewife	17	1	
	Not working	8	3	
Family history of mental illness	There	4	1	0.423
	None	51	31	
Smoking status	Yes	37	13	0.015
	None	18	19	
GRP		22.83±24.18	41.99±30.39	0.004
NMB		1.83±2.86	3.51±3.87	0.037
ALT		26.07±11.28	25.53±11.75	0.834
AST		17.00±7.03	14.50±3.98	0.037
GGT		21.96±11.29	22.90±15.78	0.768
TG		115.60±53.10	102.97±44.17	0.237
HDL		42.43±10.35	43.22±10.96	0.743
LDL		112.69±37.32	120.63±33.70	0.312
Cholesterol		161.95±29.87	172.50±32.05	0.134
FBG		80.95±12.15	83.62±10.29	0.277
HbA1c		4.91±0.57	5.23±0.42	0.004
CRP		2.96±2.09	2.83±1.50	0.729

GRP: gastrin-releasing peptide; NMB: neuromedin B; ALT: alanine aminotransferase; AST: aspartate aminotransferase; FBG: fasting blood glucose; HbA1c: hemoglobin A1c; HDL: high-density lipoprotein; LDL: low-density lipoprotein; TG: triglyceride; ADHD: attention-deficit hyperactivity disorder; SD: standard deviation; GGT: gamma-glutamyl transferase; CRP: C-reactive protein; ROC: receiver operating characteristic.

In this subscale, GRP was a strong predictor (β=-0.931, p=0.001), while the effect of NMB was not significant.

In the model created for hyperactivity/impulsivity, the first block explained 29.6% of the variance in symptoms. Adding GRP and NMB in the second block led to a significant improvement in the model, and the total variance increased to 41.3%. However, in this subscale, the independent effects of both GRP and NMB were not statistically significant.

The performance of GRP and NMB levels in distinguishing ADHD from the control group was evaluated using ROC analysis (Figure 1). The optimal cutoff value for GRP was 17.6 pg/mL, showing 84% sensitivity and 65% specificity. The optimal cutoff value for NMB was determined to be 0.905 pg/mL, providing 78% sensitivity and 64% specificity. Both biomarkers were found to have significant diagnostic power in distinguishing ADHD diagnoses.

DISCUSSION

This study found that GRP and NMB levels were significantly lower in untreated adult ADHD patients compared to healthy controls. Both biomarkers were negatively associated with ADHD symptom severity, particularly ASRS total and subscale scores, with symptoms of inattention and hyperactivity/impulsivity increasing as levels decreased. Hierarchical regression analyses showed that GRP was an independent predictor of ADHD symptom severity even after controlling for demographic and metabolic variables. ROC analyses showed that both GRP and NMB demonstrated significant discriminatory performance in distinguishing ADHD from healthy controls.

Taken together, these findings support the hypothesis that bombesin-like peptides may contribute to the biological heterogeneity of adult ADHD beyond traditional catecholaminergic mechanisms.

To our knowledge, this is the first clinical study to jointly evaluate circulating GRP and NMB levels in untreated adult ADHD patients, providing novel insights into the metabolic and neuropeptidergic underpinnings of ADHD.

The rationale for this study was based on the absence of clinical evidence regarding the combined evaluation of circulating GRP and NMB in untreated adults with ADHD, despite growing evidence linking bombesin-like peptides to metabolic regulation and central neurotransmitter systems. By examining these peptides together while controlling for major metabolic and inflammatory confounders, we sought to clarify whether they are associated with ADHD symptom

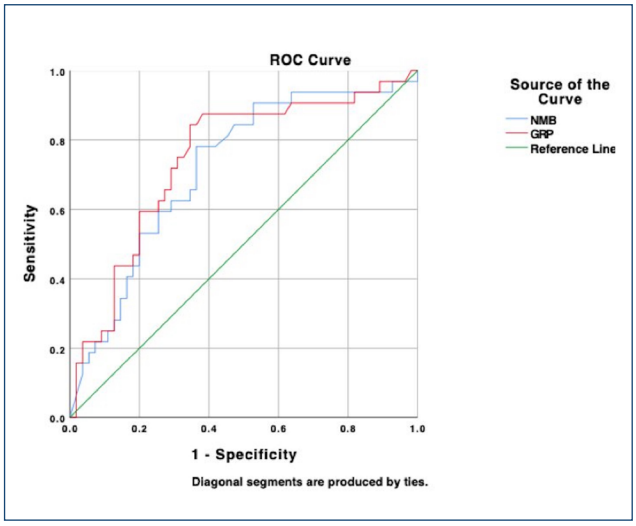


Figure 1. ROC analysis to distinguish attention-deficit hyperactivity disorder patients from healthy controls using gastrin-releasing peptide and neuromedin B levels. GRP: 95%CI 0.636–0.856, area: 0.746, standard error: 0.056, p<0.001. The optimal cutoff value determined by the maximum Youden index was 17.6, yielding a sensitivity of 84.4% and a specificity of 65.5%. NMB: 95%CI 0.608–0.830, area: 0.719, standard error: 0.057, p=0.001. The optimal cutoff value was 0.905, providing a sensitivity of 78.1% and a specificity of 63.6%.

Table 2. Hierarchical multiple regression results.

Outcome	Model	R	Adj. R^2	ΔR^2	ΔF (p)	β (GRP)	p (GRP)	β (NMB)	p (NMB)
ASRS total	Block 1	0.359	0.284	–	–	–	–	–	–
	Block 2	0.525	0.441	0.165	8,008 (0.001)	-0.642	0.024	0.071	0.777
ASRS attention	Block 1	0.342	0.265	–	–	–	–	–	–
	Block 2	0.537	0.462	0.195	9,709 (<0.001)	-0.931	0.001	-0.210	0.147
ASRS hyperactivity	Block 1	0.296	0.214	–	–	–	–	–	–
	Block 2	0.413	0.330	0.116	4,551 (0.016)	-0.276	0.426	-0.207	0.404

Notes: Block 1 covariates: Age, Sex, Smoking status, LDL, HbA1c, CRP. Block 2 predictors: GRP and NMB. ΔR^2: change in explained variance. β: standardized coefficient. Adj. R^2: adjusted R^2. ΔR^2: change in explained variance after adding Block 2 variables. β: standardized regression coefficient. Adj. R^2: adjusted R-squared. p<0.05 was considered statistically significant. NMB: neuromedin B; ASRS: Attention-Deficit Hyperactivity Disorder Self-Report Scale.

severity and whether they may represent biologically relevant markers of the disorder.

Recent studies have emphasized that ADHD is not limited to catecholaminergic dysfunction, but also involves metabolic processes and energy homeostasis. Metabolomic studies have revealed alterations in amino acid metabolism, neurotransmitter precursors, and lipid pathways, suggesting that peripheral metabolic signals may reflect central nervous system dysfunction[18,19]. In this context, the lower GRP and NMB levels observed in our study, along with their negative associations with symptom severity, are consistent with evidence indicating that peripheral metabolic dysregulation is linked to clinical manifestations of ADHD[18,19].

Evidence regarding appetite-related hormones further supports this framework. Previous studies have reported alterations in peptides such as AgRP and asprosin, as well as associations between hormones like leptin and ghrelin and ADHD symptoms[20-22]. The reduced GRP and NMB levels observed in our study align with this broader neuropeptidergic network and suggest that bombesin-like peptides may also contribute to the biological basis of ADHD.

The literature indicates that metabolic and neuropeptidergic markers may show significant associations with ADHD symptoms, but these relationships are not always maintained in multivariate models[18,19]. Consistent with this, NMB was associated with symptom severity in correlation analyses but did not emerge as an independent predictor in regression models.

In contrast, GRP remained an independent predictor even after controlling for covariates. Although GRP and NMB belong to the same peptide family, differences in their biological roles may explain this divergence. Metabolomic studies have shown that peripheral metabolites related to dopaminergic and noradrenergic pathways are associated with ADHD symptoms and executive functions[18,19,23], suggesting that peripheral signals may reflect central neurotransmitter processes. In this context, GRP may have a closer biological relationship with these pathways, which could explain its stronger association with symptom severity.

Although biomarker research in ADHD is increasing, it is widely accepted that single markers are insufficient for diagnosis in such a heterogeneous disorder. Multimodal and multimarker approaches have been shown to provide better diagnostic performance[23,24]. While GRP and NMB demonstrated significant discriminatory performance in our ROC analyses, their clinical utility as standalone diagnostic markers remains limited. Rather, these findings contribute to understanding the biological mechanisms of ADHD and may inform future biomarker-based approaches.

From a clinical perspective, GRP and NMB are not intended to replace current diagnostic procedures for ADHD. However, these biomarkers may complement future multimodal assessment strategies by providing objective biological information related to symptom severity and underlying neurobiological mechanisms. If confirmed in larger longitudinal studies, GRP in particular may help identify biologically distinct ADHD subgroups or serve as a potential biomarker for treatment monitoring and individualized therapeutic approaches.

This study has several important strengths. First, only treatment-naïve adults with ADHD were included, eliminating the potential influence of psychotropic medications on circulating peptide levels. Second, strict exclusion criteria minimized the effects of metabolic and inflammatory conditions that could affect GRP and NMB concentrations. Third, demographic, metabolic, and inflammatory variables were statistically controlled using hierarchical regression analyses, allowing the independent contribution of these neuropeptides to be evaluated. Finally, to our knowledge, this is the first clinical study to simultaneously investigate circulating GRP and NMB levels in untreated adults with ADHD.

Nevertheless, several limitations should be acknowledged. Although the sample size was sufficient according to the a priori power analysis, larger multicenter studies are needed to improve the generalizability of these findings. The cross-sectional design precludes causal inferences regarding the relationship between GRP, NMB, and ADHD symptom severity. In addition, the single-center design may limit the external validity of the findings, and residual hormonal variability in female participants cannot be completely excluded despite efforts to standardize blood sampling. Furthermore, serum GRP and NMB concentrations may not fully reflect their activity within the central nervous system; therefore, the observed associations should be interpreted cautiously until confirmed by studies integrating peripheral biomarkers with central neurobiological measures.

CONCLUSION

In conclusion, this study demonstrates that circulating GRP and NMB levels are reduced in untreated adults with ADHD and are associated with greater symptom severity. Among these biomarkers, GRP remained an independent predictor of ADHD symptom severity after adjustment for demographic, metabolic, and inflammatory factors, suggesting a potentially more prominent role in the biological mechanisms underlying the disorder. Although these findings are not sufficient to support the clinical use of GRP or NMB as standalone diagnostic biomarkers, they provide further evidence for the involvement of metabolic and

neuropeptidergic pathways in ADHD and support future multimodal biomarker research. Larger longitudinal and mechanistic studies are needed to validate these findings and determine their potential role in diagnosis, prognosis, and treatment monitoring.

AUTHORS' CONTRIBUTIONS

EÖT: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. **ALÖ:** Data curation, Investigation, Methodology, 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. Biederman J. Attention-deficit/hyperactivity disorder: a selective overview. Biol Psychiatry. 2005;57(11):1215-20.
2. Roesler R, Henriques J, Schwartsmann G. Gastrin-releasing peptide receptor as a molecular target for psychiatric and neurological disorders. CNS Neurol Disord Drug Targets. 2006;5(2):197-204.
3. Sayegh AI. The role of bombesin and bombesin-related peptides in the short-term control of food intake. Prog Mol Biol Transl Sci. 2013;114:343-70.
4. Ohki-Hamazaki H. Neuromedin B. Prog Neurobiol. 2000;62(3):297-312.
5. Paula GS, Wilieman M, Silva KR, Baptista LS, Boudina S, Souza LL, et al. Neuromedin B receptor disruption impairs adipogenesis in mice and 3T3-L1 cells. J Mol Endocrinol. 2019;63(1):93-102.
6. Hoggard N, Bashir S, Cruickshank M, Miller JD, Speakman J. Expression of neuromedin B in adipose tissue and its regulation by changes in energy balance. J Mol Endocrinol. 2007;39(3):199-210.
7. Bedard T. Characterization of the role of neuromedin B and gastrin-releasing peptide in stress and anxiety-related responses [dissertation]. Ottawa: University of Ottawa; 2006.
8. Goto F, Kiyama Y, Ogawa I, Okuno H, Ichise T, Ichise H, et al. Gastrin-releasing peptide regulates fear learning under stressed conditions via activation of the amygdalostriatal transition area. Mol Psychiatry. 2022;27(3):1694-703.
9. Becker MM, Riesgo RS, Roesler R, Bosa C, Ohlweiler L, Backes B, et al. Improvement in symptoms of autism spectrum disorder with gastrin-releasing peptide: an open trial. Clin Neuropharmacol. 2016;39(5):215-9.
10. Marchezan J, Becker M, Schwartsmann G, Ohlweiler L, Roesler R, Renck LB, et al. A placebo-controlled crossover trial of gastrin-releasing peptide in childhood autism. Clin Neuropharmacol. 2017;40(3):108-12.
11. Kabayama Y, Kato Y, Shimatsu A, Yanaihara N, Imura H. Stimulation of dopamine and norepinephrine release by gastrin-releasing peptide and related peptides. Brain Res. 1986;372(2):394-9.
12. Guo M, Qu X, Qin XQ. Bombesin-like peptides and their receptors: recent findings in pharmacology and physiology. Curr Opin Endocrinol Diabetes Obes. 2015;22(1):3-8.
13. Ter Beek W, Muller E, Hogezand R, Biemond I, Lamers C. Gastrin-releasing peptide receptor expression in inflammatory bowel disease. J Clin Pathol. 2004;57(10):1047-51.
14. Zhu T, Chen J, Zhao Y, Zhang J, Peng Q, Huang J, et al. Neuromedin B mediates inflammatory signaling pathways in human cells. Biosci Rep. 2019;39(10):BSR20192139.
15. Kooij JJS. Adult ADHD: diagnostic assessment and treatment. New York: Springer; 2012.
16. Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi E, et al. The World Health Organization adult ADHD self-report scale (ASRS). Psychol Med. 2005;35(2):245-56.
17. Doğan S, Öncü Çetinkaya B, Saraçoğlu Varol G, Küçükgöncü S. Validity and reliability of the Turkish version of ASRS. Anadolu Psikiyatri Derg. 2009;10:77-87.
18. Predescu E, Vaidean T, Rapciuc AM, Sipos R. Metabolomic markers in ADHD: a systematic review. Int J Mol Sci. 2024;25(8):4385.
19. Muñoz-Zabaleta MJ, Garzón Rodríguez N, Díaz-Barrera LE, Quiroz-Padilla MF. Metabolomic insights into ADHD: a scoping review. Metabolites. 2025;15(2):133.
20. Gürbüzer N, Zengil S, Laloğlu E, Kaşali K. Appetite-related peptides and metabolic parameters in ADHD. Arch Neuropsychiatry. 2024;61(1):30-8.
21. Marcelli I, Capece U, Caturano A. ADHD and metabolic disorders: shared mechanisms. Metabolites. 2025;15(5):40.
22. Lai KY, Li CJ, Tsai CS, Chou WJ, Huang WT, You HL, et al. Appetite hormones and ADHD: associations with orexin, leptin and ghrelin. Psychopharmacology (Berl). 2024;241(2):91-102.
23. Hung YA, Kuo TC, Tseng YJ, Shang CY, Gau SSF. Novel metabolites in ADHD identified by metabolome profiling. Transl Psychiatry. 2025;15(1):180.
24. Hurjui IA, Hurjui RM, Hurjui LL, Serban IL, Dobrin I, Apostu M, et al. Biomarkers and neuropsychological tools in ADHD. Medicina (Kaunas). 2025;61(7):1211.

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260669

The effect of micronutrient supplementation on women's sexual desire: a systematic review and meta-analysis study

Nurçin Küçük Kent[1], Sevda Karakaş[2]*, Sevda Uzun[3]

SUMMARY

OBJECTIVE: The aim of this study was to determine the effectiveness of micronutrient supplementation on female sexual desire levels.

METHODS: For this systematic review and meta-analysis, studies were identified by searching the PubMed, Web of Science, ScienceDirect, Google Scholar, and Higher Education Council (YÖK) Thesis Center databases without any year restrictions. After screening, eight studies were included in the analysis. Data were synthesized using meta-analysis and narrative methods.

RESULTS: In this systematic review and meta-analysis, micronutrient supplementation was associated with a statistically significant effect on women's sexual desire (standardized mean difference: -0.597, 95%CI -1.126 to -0.068, Z=-2.211, p=0.027); however, substantial heterogeneity was observed (I^2 =87.2%) indicating considerable variability across studies.

CONCLUSION: This systematic review suggests a potential role for micronutrients, but the evidence is weak and highly inconsistent. High-quality, low-heterogeneity research is urgently needed.

KEYWORDS: Micronutrients. Women. Sexual desire. Meta-analysis.

INTRODUCTION

Micronutrients (vitamins and minerals) contribute to maintaining the physiological homeostasis of sexual desire and sexual function in women and to preventing the development of potential sexual dysfunction by playing a role in the regulation of gonadal functions and in the steroidogenesis process of sex-specific hormones[1,2]. Micronutrients, particularly zinc, vitamin D, B vitamins, and antioxidant vitamins, play a supportive role in maintaining women's sexual health and improving potential sexual dysfunction through their effects on hormonal balance, neurotransmitter activity, and vascular function[3,4]. It is known that micronutrient deficiency is a significant risk factor for sexual aversion in women. Sexual aversion, one of the most common issues among sexual dysfunctions in women, ranks among the most prevalent disorders, with a prevalence rate of 50.7%[5,6]. Studies in the literature report that micronutrient supplementation (zinc, vitamin D, and B vitamins) provides significant benefits in maintaining women's sexual health and improving sexual desire by supporting hormone synthesis, neurotransmitter balance, and overall metabolic functions[6-8]. Zinc may contribute to the regulation of sexual desire by playing a role in the synthesis of gonadal hormones, including estrogen and progesterone. The presence of vitamin D receptors in the reproductive system and brain tissue suggests that this vitamin may influence libido through its effects on sexual function and mood. It has been suggested that B vitamins may indirectly increase sexual desire and arousal by supporting neurotransmitter synthesis through serotonergic and dopaminergic pathways[8,9]. This study was conducted to investigate the effectiveness of micronutrient supplementation on women's sexual desire levels.

METHODS

This systematic review and meta-analysis was prepared in accordance with the PRISMA checklist (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols)[10]. To minimize the risk of potential bias in this systematic review and meta-analysis, the literature search, article selection, and data extraction processes were conducted independently by

[1]Gümüşhane University, Department of Emergency Aid and Disaster Management – Gümüşhane, Türkiye.
[2]Gümüşhane University, Faculty of Health Sciences, Department of Nursing – Gümüşhane, Türkiye.
[3]Kırşehir Ahi Evran University, Faculty of Health Sciences, Department of Nursing – Kırşehir, Türkiye.
*Corresponding author: sevdakarakas@gumushane.edu.tr
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on March 25, 2026. Accepted on August 01, 2026.
Scientific Editor: José Maria Soares Júnior

three researchers. These steps were subsequently reviewed by the three researchers. The quality assessment of the studies included in the systematic review and meta-analysis was also performed by the researchers.

Inclusion criteria

In this study, studies were screened according to PICOS:

- P: Population: Women
- I: Intervention: Micronutrient use
- C: Comparison: No use of micronutrients
- C: Outcomes: Sexual desire
- S: Study design: Experimental studies published in Turkish and English.

Exclusion criteria

Studies in the form of letters to the editor, case–control studies, and systematic and traditional reviews were excluded from the scope of this research.

Search strategy

The search was conducted through ScienceDirect, Web of Science, YÖK TEZ, and Google Scholar up to March 2026, using the keywords "sexual desire and women," "micronutrients and women and sexual desire," or "micronutrients and women," in accordance with Medical Subject Headings (MeSH) criteria. A search was conducted on PubMed using appropriate MESH terms ("Sexual Desire"[MESH] OR "Sexual Dysfunction, Physiological"[MESH]) AND ("Female"[MESH]) AND ("Micronutrients"[Mesh] OR "Vitamins"[Mesh] OR "Minerals"[Mesh] OR "Dietary Supplements"[Mesh]). Since the number of studies on the effects of micronutrient supplementation on women's levels of sexual desire is limited, no year restrictions were applied, and all years were included in the search.

Selection of studies

The search initially yielded 146 records. After removing duplicate studies, 131 records were reviewed based on their titles and abstracts for selection. As a result of this review, 48 studies were selected for full-text analysis. Subsequently, the 31 articles for which full texts were obtained were reviewed according to inclusion and exclusion criteria, and 8 studies reporting results on the effects of micronutrient supplementation on women's sexual desire levels were included in the analysis[11-18]. Explanations regarding the article selection process are provided in Figure 1. This study is a systematic review and meta-analysis and was conducted based on studies published in the literature.

Assessment of the methodological quality of the studies

The quality assessment of the studies included in this systematic review and meta-analysis was conducted by the Joanna Briggs Institute using quality assessment tools tailored to the study design. The assessment tools used in this study were selected based on the designs of the studies included in the systematic review and meta-analysis. In our study, a 13-item assessment tool was used for randomized controlled studies (The Joanna Briggs Institute Critical Appraisal Tools for Use in IBI Systematic Reviews, 2021), and a 9-item assessment tool was used for quasi-experimental studies. The methodological quality of the included studies was categorized based on the percentage of "Yes" responses to the 13 JBI criteria: ≥70% was considered low risk of bias (high quality), 50–69% as moderate risk of bias (moderate quality), and <50% as high risk of bias (low quality). No studies were found to have a high risk of bias. In this study, the evaluation results for each study are shown in Table 1 as "*Quality score*"[19].

Data synthesis

For the statistical calculations of this study, CMA Ver. 2 was used. The heterogeneity between the studies was evaluated with the Chi-squared statistic and Higgins I^2 tests, and an I^2 of more than 50% was considered to indicate significant heterogeneity[19]. Studies with $I^2 \leq 50\%$ were evaluated using a fixed effects model; however, studies with $I^2 > 50\%$ were evaluated using a random effects model. A random effects model was used in this study. Statistical significance was considered at $p<0.05$ for all tests[19,20].

In addition, the Tau-squared statistic was used to complete the assessment of variance and heterogeneity between studies. The standardized mean difference (SMD) with 95%CI was used to assess the effect size associated with the same outcome with different measurement tools. Furthermore, forest plots were prepared to visualize the SMD with 95%CI. These D values were then averaged to calculate the overall effect size. The D value was converted to a Z value to assess statistical significance. Funnel plots were used to examine and visualize publication bias. Publication bias was detected by an asymmetry in the funnel plot. The Begg test was also performed to assess publication bias[19,20].

Sensitivity analysis was performed by excluding one study at a time and evaluating whether there was a significant change in the results; there was no significant change in the results[19,20].

RESULTS

The selection of studies for this systematic review and meta-analysis according to the PRISMA flow diagram is presented in Figure 1.

The characteristics and main findings of the studies are presented in Table 1; it was determined that the studies met more than half of the items in the evidence quality assessment tool (Table 1).

Meta-analysis results on the effectiveness of micronutrient supplementation regarding women's sexual desire level

In this study, the presence of publication bias was determined using two methods: (a) a funnel plot and (b) the Begg and Mazumdar correlation test. In the funnel plot—one of the key methods for detecting publication bias—we observe that the studies in this dataset are clustered at the top of the funnel and exhibit a symmetrical distribution. This indicates the absence of publication bias (Figure 2).

Publication bias among the studies included in this dataset was determined using the Begg and Mazumdar rank correlation method. According to the rank correlation method, the two-sided p-value is 0.22. This result indicates that the publication bias is not statistically significant.

Figure 3 presents a forest plot of the eight studies examining the effect of micronutrient supplementation on women's sexual desire levels, including effect sizes, standard errors, variances, and lower and upper limits. Based on the findings of these studies, a meta-analysis concluded that micronutrient supplementation is effective in improving women's sexual desire levels (SMD: -0.597, 95%CI -1.126 to -0.068; Z=-2.211, p=0.027, I^2=87.2%) (Figure 3).

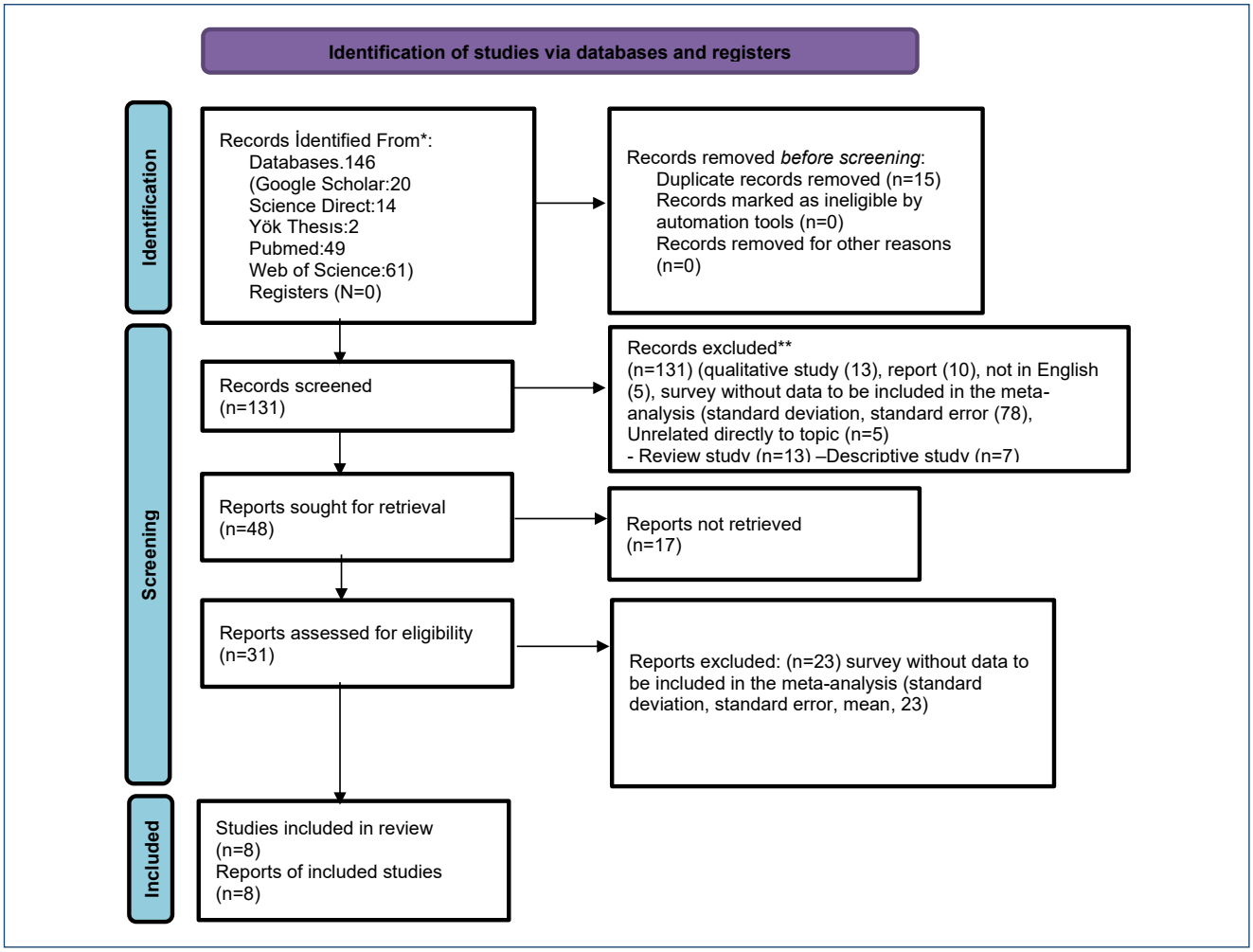


Figure 1. PRISMA. *Databases searched: Google Scholar, ScienceDirect, YÖK Thesis, PubMed, and Web of Science. **Reasons for exclusion: qualitative studies, reports, studies not published in English, surveys without sufficient data for meta-analysis, studies not directly related to the topic, review studies, and descriptive studies.

DISCUSSION

This study was conducted to systematically review randomized controlled trials evaluating the effectiveness of micronutrient supplementation on female sexual desire and to perform a meta-analysis of the existing evidence. The findings of this systematic review and meta-analysis suggest that micronutrient supplementation may improve female sexual desire; however, the certainty of the available evidence is limited. Similar

Table 1. Characteristics and main findings of the studies included in the systematic review and meta-analysis.

Author/year	Study design	Measurement tools	Sample size/ characteristics	Duration of intervention	Food used in the intervention	Main results	Quality scores
Izadi et al. [11]	Randomized controlled trial	Female Sexual Function Index (FSFI)	Experimental group: 25 Control group: 25	8 weeks	Saffron	An increase in sexual desire was observed in the experimental group.	Yes: 13/13
Valizadeh et al.[13]	Randomized controlled trial	Female Sexual Function Index (FSFI)	Experimental group: 30 Control group: 29	8 weeks	Probiotic	An increase in sexual desire was observed in the experimental group.	Yes: 13/13
Mazaheri Nia et al.[10]	Randomized controlled trial	Female Sexual Function Index (FSFI)	Experimental group: 30 Control group: 29	8 weeks	Zinc	An increase in sexual desire was observed in the experimental group.	Yes: 9/13 No: 2/13 Uncertain: 1/13 Not applicable: 1/13
Ghamari et al.[12]	Randomized controlled trial	Female Sexual Function Index (FSFI)	Experimental group: 31 Control group: 31	8 weeks	Ginseng	An increase in sexual desire was observed in the experimental group.	Yes: 13/13
Jahanian Sadatmahalleh et al.[14]	Randomized controlled trial	Female Sexual Function Index (FSFI)	Experimental group: 35 Control group: 35	8 weeks	Vitamin D	An increase in sexual desire was observed in the experimental group.	Yes: 13/13
Sarebani et al.[17]	Randomized controlled trial	Female Sexual Function Index (FSFI)	Experimental group: 33 Control group: 32	8 weeks	Vitamin D	An increase in sexual desire was observed in the experimental group.	Yes: 9/13 No: 2/13 Uncertain: 1/13 Not applicable: 1/13
Rafati et al.[15]	Randomized controlled trial	Female Sexual Function Index (FSFI)	Experimental group: 20 Control group: 20	4 weeks	Vitamin D	An increase in sexual desire was observed in the experimental group.	Yes: 13/13
Ito et al.[16]	Randomized controlled trial	Female Sexual Function Index (FSFI)	Experimental group: 34 Control group: 43	8 weeks	Arginmax	An increase in sexual desire was observed in the experimental group.	Yes: 13/13

to our study, Canat et al. found in their study involving 50 women with low sexual desire that micronutrient supplementation (vitamin D) was effective in increasing women's sexual desire levels[3]. The effects of vitamins A, C, E, D, and B on women's sexual health have been examined in the literature; it has been reported that vitamins A, C, E, D, and B complex vitamins have beneficial effects on women's sexual function[8-12]. In particular, study results have reported that vitamin D and niacin (B3) have effects that increase women's sexual desire[9-11]. The pooled findings suggest that micronutrient supplementation may have a beneficial effect on women's sexual desire; however, this finding should be interpreted with caution due to the substantial heterogeneity and variability in study designs. Female sexual desire disorder is a multifactorial condition, and therefore a single standardized treatment approach is not appropriate. Existing literature suggests that micronutrient supplementation may positively influence female sexual function through potential mechanisms such as supporting hormonal balance, reducing oxidative stress, regulating neurotransmitter synthesis, and improving vascular function[11,21,22]. The management of female sexual desire disorder requires a multidisciplinary approach, typically beginning with individualized psychotherapy and, when appropriate, proceeding to pharmacological treatments and, in rare cases, surgical interventions. Within this framework, existing literature suggests that micronutrient supplementation may positively influence female sexual function through potential mechanisms such as supporting hormonal balance, reducing oxidative stress, regulating neurotransmitter synthesis, and improving vascular function; therefore, adequate vitamin and mineral intake may have a supportive role in maintaining sexual function[12-16]. In contrast to our study, Ghanbari et al. reported in their 2023 study involving 337 women that vitamin supplementation (including vitamins A, C, E, and B) did not result in a significant improvement in female sexual function, that the results were inconsistent, and that the overall level of evidence was insufficient[23]. It is believed that the discrepancy in results is related to differences in intervention procedures and inclusion criteria. Therefore, although this systematic review suggests a potential role for micronutrient supplementation in improving

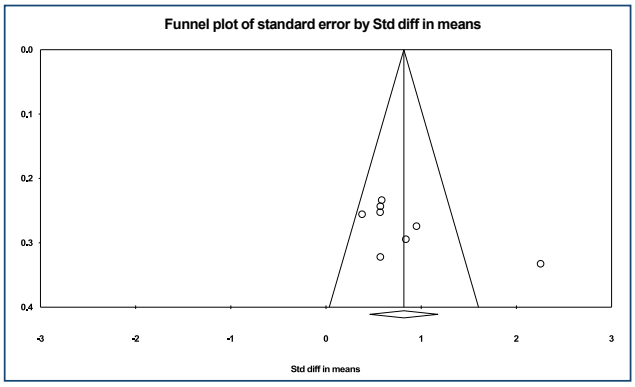


Figure 2. Funnel plot of studies reporting results on the effectiveness of micronutrient supplementation on women's sexual desire levels.

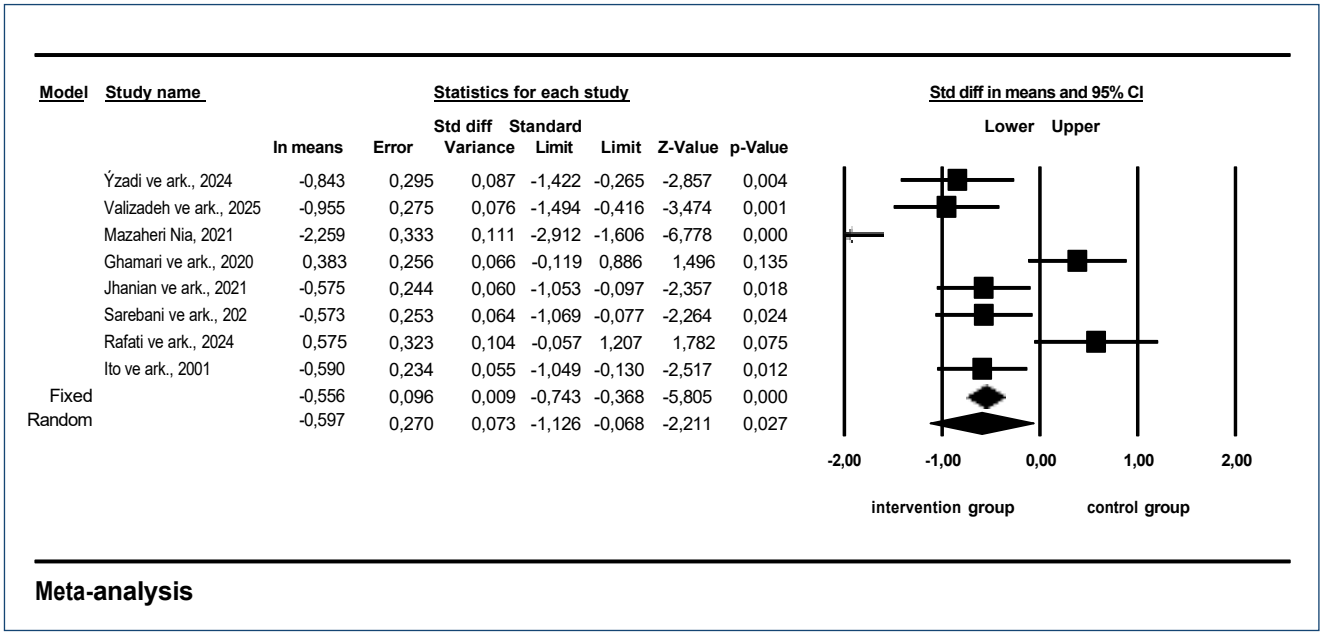


Figure 3. Forest plot showing the effect of micronutrient supplementation on women's sexual desire levels in the intervention and control groups.

female sexual desire, the available evidence remains weak and highly heterogeneous. Future well-designed randomized controlled trials with larger sample sizes, standardized intervention protocols, and validated outcome measures are urgently needed to provide more robust evidence.

CONCLUSION AND RECOMMENDATIONS

This systematic review and meta-analysis suggest that micronutrient supplementation may improve women's sexual desire; however, the current evidence is weak and highly inconsistent, and the certainty of the available evidence is limited. Additionally, it is recommended that randomized controlled trials be conducted with larger sample sizes to evaluate the effectiveness of micronutrient supplementation on women in terms of various variables. Future studies should also consider evaluating additional psychosocial and biological variables that may influence women's sexual desire in order to provide a more comprehensive understanding of the effects of micronutrient supplementation.

LIMITATIONS OF THE STUDIES

Some of the studies included in the systematic review and meta-analysis had small sample sizes and were conducted without blinding. These factors may reduce the strength of the evidence provided by the studies.

AUTHORS' CONTRIBUTIONS

NKK: Data curation, Formal Analysis, Investigation, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **SK:** Conceptualization, Data curation, Formal Analysis, Investigation, Software, Supervision, Validation, Visualization, Writing – original draft; Writing – review & editing. **SU:** Data curation, Supervision, 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. Sarmento ACA, Costa APF, Lírio J, Eleutério J Jr, Baptista PV, Gonçalves AK. Efficacy of hormonal and nonhormonal approaches to vaginal atrophy and sexual dysfunctions in postmenopausal women: a systematic review. Rev Bras Ginecol Obstet. 2022;44(10):986-94. https://doi.org/10.1055/s-0042-1756148
2. Pilz S, Frisch S, Koertke H, Kuhn J, Dreier J, Obermayer-Pietsch B, et al. Vitamin D and endocrine regulation: implications for human health. Horm Metab Res. 2011;43(3):223-5. https://doi.org/10.1055/s-0030-1269854
3. Canat M, Canat L, Öztürk FY, Eroğlu H, Atalay HA, Altuntaş Y. Vitamin D_3 deficiency is associated with female sexual dysfunction in premenopausal women. Int Urol Nephrol. 2016;48:1789-95. https://doi.org/10.1007/s11255-016-1396-7
4. Davis SR, Melhorn SJ. Vitamin A and reproductive health: a review. Endocr Rev. 2018;39(4):465-82.
5. Alidost F, Pakzad R, Dolatian M, Abdi F. Sexual dysfunction among women of reproductive age: a systematic review and meta-analysis. Int J Reprod Biomed. 2021;19(5):421-32. https://doi.org/10.18502/ijrm.v19i5.9251
6. Zhang H, Yip PSF. Prevalence and correlates of female sexual dysfunction and sexual distress in reproductive-aged women: a systematic review and meta-analysis. BMC Womens Health. 2025;25:39. https://doi.org/10.1186/s12905-025-03960-4
7. Prasad AS. Zinc: an antioxidant and anti-inflammatory agent and its role in human health. J Trace Elem Med Biol. 2013;27(2):75-82. https://doi.org/10.1016/j.jtemb.2012.10.001
8. Smith AD, Refsum H. Homocysteine, B vitamins, and cognitive impairment. Ann Rev Nutr. 2016;36:211-39. https://doi.org/10.1146/annurev-nutr-071715-050947
9. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for systematic reviews. BMJ. 2021;372:n71. https://doi.org/10.1136/bmj.n71
10. Mazaheri Nia L, Iravani M, Abedi P, Cheraghian B. Effect of zinc on testosterone levels and sexual function of postmenopausal women: a randomized controlled trial. J Sex Marital Ther. 2021;47(8):804-13. https://doi.org/10.1080/0092623X.2021.1957732
11. Izadi S, Mohammad-Alizadeh-Charandabi S, Yadollahi P, Mirghafourvand M. Effect of vitamin E with and without saffron on sexual function in women with sexual dysfunction: a randomized controlled trial. BMC Womens Health. 2024;24:143. https://doi.org/10.1186/s12905-024-02980-w
12. Ghamari K, Kashani L, Jafarinia M, Najafabadi BT, Shokraee K, Esalatmanesh S, et al. Vitamin E and ginseng supplementation to enhance female sexual function: a randomized, double-blind, placebo-controlled, clinical trial. Women Health. 2020;60(10):1164-73. https://doi.org/10.1080/03630242.2020.1803465
13. Valizadeh R, Borhani F, Shabani F, Mirghafourvand M. The effect of probiotic supplementation on sexual function in women: a randomized controlled trial. Sci Rep. 2025;16(1):1254. https://doi.org/10.1038/s41598-025-30879-5
14. Jahanian Sadatmahalleh S, Razavinia F, Lotfi R, Hajimoosayi F, Nasiri M, Rezasoltani P. The effect of hormonal contraceptives combined with vitamin D3 supplements on sexual dysfunction in women with polycystic ovary syndrome: a randomized double-blind placebo-controlled trial. Arch Gynecol Obstet. 2025;312(5):1737-46. https://doi.org/10.1007/s00404-025-08159-4
15. Rafati M, Bazrafshan E, Shaki F, Ghalini-Moghaddam T, Moghimi M. The relationship between serum vitamin D, testosterone, and

oxidative stress levels in women with sexual dysfunction: a case-controlled study. Taiwan J Obstet Gynecol. 2024;63(5):673-8. https://doi.org/10.1016/j.tjog.2024.06.004

16. Ito TY, Trant AS, Polan ML. ArginMax and female sexual function: a randomized placebo-controlled study. J Sex Marital Ther. 2001;27(5):541-9. https://doi.org/10.1080/009262301317081070
17. Sarebani Z, Chegini V, Chen H, Aali E, Mirzadeh M, Abbaspour M, et al. Effect of vitamin D vaginal suppository on sexual functioning among postmenopausal women: a three-arm randomized controlled clinical trial. Obstet Gynecol Sci. 2023;66(3):208-20. https://doi.org/10.5468/ogs.22038
18. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to meta-analysis. Wiley; 2021.
19. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-60. https://doi.org/10.1136/bmj.327.7414.557
20. The Joanna Briggs Institute. Critical appraisal tools for use in JBI systematic reviews [Internet]. 2017. [cited on 2026 May 01]. Available from: http://joannabriggs.org/research/critical-appraisal-tools.html
21. Sha'ari N, Woon LSC, Sidi H, Das S, Bousman CA, Saini SM. Beneficial effects of natural products on female sexual dysfunction: a systematic review and meta-analysis. Phytomedicine. 2021;93:153760. https://doi.org/10.1016/j.phymed.2021.153760
22. Quesada S, Furlan V, Meling A, Hammad M, Yih J. Exploring vitamin D as a modifiable factor in female sexual dysfunction: a systematic review. J Sex Med. 2026;23(Suppl 3):qdag063.105. https://doi.org/10.1093/jsxmed/qdag063.105
23. Ghanbari-Homaie S, Ataei-Almanghadim K, Mirghafourvand M. Effect of vitamins on sexual function: a systematic review. Int J Vitam Nutr Res. 2023;93(2):154-63. https://doi.org/10.1024/0300-9831/a000703

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260806

Partner hygiene–related behaviors and vaginitis subtypes: an exploratory observational study

Duygu Ucar Kartal[1]*, Mehmet Emre Peker[1], Ufuk Atlihan[2], Murat Ozsahin[3], Haydar Kaya[1]

SUMMARY

OBJECTIVE: The aim of this study was to investigate the association between partner hygiene behaviors, quantified using an exploratory partner hygiene score, and vaginitis subtypes, and to evaluate the relationship between baseline partner hygiene and short-term recurrence.

METHODS: This prospective observational study included women aged 18–45 years presenting with vaginitis symptoms to a gynecology outpatient clinic. Vaginitis subtypes were diagnosed using standardized clinical and laboratory criteria. Partner hygiene behaviors were assessed using a seven-item partner hygiene score (range: 0–14). Participants were followed for 3 months to assess clinically confirmed symptomatic recurrence. Associations between partner hygiene score and vaginitis subtypes were analyzed cross-sectionally, while recurrence analyses were performed prospectively.

RESULTS: A total of 108 women were included. Diagnoses included bacterial vaginosis (45.4%), vulvovaginal candidiasis (29.6%), ***Trichomonas vaginalis*** infection (10.2%), mixed vaginitis (11.1%), and non-specific vaginitis (3.7%). Median partner hygiene score scores were significantly lower in women with bacterial vaginosis and trichomoniasis compared with vulvovaginal candidiasis ($p<0.001$). Each one-point decrease in partner hygiene score was associated with higher odds of bacterial vaginosis. During follow-up, recurrence occurred in 33% of cases, but baseline partner hygiene score was not associated with recurrence.

CONCLUSION: Lower partner hygiene scores were associated with dysbiosis-predominant vaginitis phenotypes, particularly bacterial vaginosis and trichomoniasis. Although no association was observed with short-term recurrence, these findings are hypothesis-generating and require confirmation in larger prospective studies.

KEYWORDS: Vaginitis. Hygiene. Vaginosis, bacterial. Candidiasis, vulvovaginal. Trichomonas vaginitis.

INTRODUCTION

Vaginal infections—particularly bacterial vaginosis (BV), vulvovaginal candidiasis, and *Trichomonas vaginalis* infection—represent a major clinical and public health concern worldwide, with BV affecting approximately 23–29% of women of reproductive age globally, vulvovaginal candidiasis affecting up to 75% of women during their lifetime, and *T. vaginalis* infection accounting for an estimated 156 million new cases annually[1-4]. Central to these conditions is disruption of the *Lactobacillus*-dominated vaginal ecosystem, which predisposes to dysbiosis, polymicrobial biofilm formation, and increased susceptibility to sexually transmitted infections and poor pregnancy outcomes[5].

Despite extensive research on host-related determinants, emerging evidence indicates that vaginal infections cannot be fully understood without accounting for partner-level contributors. Penile microbiota composition, particularly anaerobe-enriched communities, has been strongly linked to incident BV in female partners, while male circumcision modifies this microbial ecology and may reduce BV and HIV risk[6,7]. Yet findings remain inconsistent, underscoring the complexity of partner influences[8].

Partner-related behavioral and hygiene factors may influence vaginal infection susceptibility; however, these variables have rarely been assessed using a standardized quantitative framework[9-17]. The lack of a standardized assessment of partner hygiene remains an important gap in vaginitis research[14,17,18].

Therefore, this study aimed to explore the cross-sectional relationship between partner hygiene–related behaviors and common vaginitis subtypes among symptomatic women attending a gynecology outpatient clinic. Secondary analyses explored associations between baseline partner hygiene and 3-month recurrence.

[1]Manisa Merkezefendi State Hospital, Department of Obstetrics and Gynecology – Manisa, Turkey.
[2]Izmir Democracy University, Buca Seyfi Demirsoy Training and Research Hospital, Department of Obstetrics and Gynecology – İzmir, Turkey.
[3]Manisa City Hospital, Department of Obstetrics and Gynecology – Manisa, Turkey.
*Corresponding author: drdygucar@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on June 30, 2026. Accepted on August 04, 2026.
Scientific Editor: José Maria Soares Júnior

METHODS

Women aged 18–45 years who presented with complaints of vaginal discharge, odor, and/or itching, reported a history of sexual intercourse within the preceding 3 months, and were clinically diagnosed with vaginitis were consecutively included in the study. Pregnant women, those who had used antibiotics within the previous 4 weeks, women with known immunodeficiency, and those without a sexual partner were excluded. Women older than 45 years and those with postmenopausal status were excluded to minimize the potential influence of menopausal hormonal changes on the vaginal microbiota were excluded to minimize the potential influence of perimenopausal hormonal changes on the vaginal microbiota. Women with known immunodeficiency were excluded because of their increased susceptibility to atypical and recurrent vaginal infections, which could confound the association between partner hygiene and vaginitis subtypes.

All participants provided written informed consent prior to enrollment. The study protocol was approved by the Ethics Committee of the Istanbul Esenyurt University (Approval No. 2025/10-21) and was conducted in accordance with the principles of the Declaration of Helsinki.

Participants completed a structured questionnaire covering sociodemographic characteristics, reproductive history, sexual practices, personal hygiene habits, and previous history of vaginal infections.

Clinical diagnosis of vaginitis was established by standardized gynecologic examination supported by laboratory findings, including vaginal pH assessment, microscopy, and culture where indicated. BV was diagnosed using the Amsel criteria (≥3 of 4 criteria) and/or a Nugent score of 7–10. Vulvovaginal candidiasis was diagnosed based on compatible clinical findings together with microscopy and/or positive fungal culture. Mixed vaginitis was defined as the simultaneous presence of diagnostic criteria for two or more vaginal infections. Non-specific vaginitis was defined as symptomatic vaginitis in the absence of diagnostic criteria for BV, vulvovaginal candidiasis, or *T. vaginalis* infection. All clinical examinations and diagnostic assessments were performed by a single experienced gynecologist according to predefined standardized diagnostic criteria. Therefore, inter-rater reliability was not applicable.

Treatment

All participants received treatment according to current international guidelines and the diagnosed vaginitis subtype. Women with BV received oral metronidazole (500 mg twice daily for 7 days). Vulvovaginal candidiasis was treated with a single 150-mg oral dose of fluconazole or topical azole therapy, as clinically indicated. Patients with *T. vaginalis* infection received oral metronidazole (500 mg twice daily for 7 days), and partner treatment was recommended according to current guidelines. Mixed vaginitis was treated according to the identified pathogens.

Partner hygiene score

Partner hygiene behaviors were quantitatively assessed using the partner hygiene score (PHS), a composite score consisting of seven equally weighted items evaluating cleansing frequency, cleansing product type, underwear change frequency, pre- and post-coital hygiene, genital hair management–related skin integrity, and recurrent hygiene-related genital irritation. The items were selected based on their biological plausibility and clinical relevance to partner hygiene practices potentially influencing the vaginal microenvironment. Each item was scored from 0 to 2, yielding a total score ranging from 0 to 14, with higher scores indicating better partner hygiene. The PHS items were developed based on clinically relevant partner hygiene behaviors considered potentially associated with vaginal health, and the rationale for item selection was based on current clinical knowledge and literature findings.

The total PHS score was calculated and analyzed as a continuous variable, with higher scores indicating better partner hygiene. The total PHS score ranged from 0 to 14 and was analyzed both as a continuous variable and by tertiles. Internal consistency of the seven-item PHS was assessed using Cronbach's alpha, demonstrating acceptable internal consistency (Cronbach's α=0.793). The complete PHS questionnaire, item descriptions, and scoring instructions are provided in Supplementary Appendix 1.

Follow-up and outcome measures

Participants were followed prospectively for 3 months after initial diagnosis and treatment. Treatment adherence was assessed at the 3-month follow-up visit by patient self-report. Women who reported that they had not completed the prescribed treatment regimen (n=5) were excluded from the recurrence analysis. Consequently, recurrence analyses were performed in 103 women. At the 3-month follow-up visit, all participants underwent clinical reassessment, and recurrent vaginitis was diagnosed using the same predefined diagnostic criteria applied at baseline. Recurrence was defined as the reappearance of symptomatic vaginitis confirmed by clinical and laboratory evaluation during the follow-up period.

The primary outcome:

- Vaginitis subtype at baseline (BV, VVC, TV, mixed, non-specific).

Secondary outcome (exploratory):

- Symptomatic recurrence within 3 months after treatment, confirmed by clinical/laboratory evaluation.

As this was an exploratory observational study, no formal a priori sample size calculation was performed. Statistical analyses were performed using SPSS version 24.0. Continuous variables were presented as mean±standard deviation or median (interquartile range), as appropriate. Categorical variables were summarized as frequencies and percentages. Comparisons between groups were performed using Student's t-test or Mann-Whitney U test for continuous variables and χ^2 test or Fisher's exact test, as appropriate, for categorical variables. Multivariable logistic regression models were constructed to assess the independent association between the PHS score and BV. To minimize the risk of model overfitting given the available sample size, variables were selected a priori based on established clinical relevance and evidence from previous literature. The final adjusted model included age, body mass index, smoking status, number of sexual partners during the previous 12 months, vaginal douching, and PHS score. The variable of having ≥2 sexual partners during the previous 12 months was retained in the multivariable model based on its established relevance as a potential behavioral confounder, despite not reaching statistical significance in univariable analysis. Adjusted odds ratios (aORs) with 95%CIs were calculated. A two-sided $p<0.05$ was considered statistically significant. Given the exploratory nature of the study, p-values were not adjusted for multiple comparisons and should be interpreted as exploratory. For categorical comparisons involving PHS tertiles, χ^2 test results were interpreted cautiously because of small cell sizes in some subgroup comparisons. A post-hoc sensitivity analysis was performed using G*Power version 3.1 (Heinrich Heine University, Düsseldorf, Germany). For the recurrence analysis (n=103), the minimum detectable effect size was Cohen's d=0.593 (α=0.05, power=80%).

RESULTS

A total of 152 women were assessed for eligibility during the study period. After applying the inclusion and exclusion criteria, 131 participants were enrolled. A total of 23 participants were excluded from the final analysis due to loss to follow-up or incomplete questionnaires and clinical data, resulting in a final study population of 108 women. The baseline demographic and clinical characteristics of the study population are presented in Table 1. The mean age of the participants was 31.1±5.8 years, and all reported sexual intercourse within

Table 1. Baseline demographic, behavioral, and clinical characteristics of the study population (n=108).

Characteristic	Value
Demographic characteristics	
Age (years), mean±SD	31.1±5.8
Body mass index (kg/m²), mean±SD	26.1±4.3
Higher education*, n (%)	58 (53.7)
Middle/high socioeconomic status†, n (%)	66 (61.1)
Current smoking, n (%)	31 (28.7)
Alcohol use‡, n (%)	30 (27.8)
Reproductive and sexual characteristics	
Gravidity, median (IQR)	2 (1–3)
Parity, median (IQR)	1 (1–2)
Nulliparous, n (%)	32 (29.6)
Previous vaginal delivery, n (%)	58 (53.7)
Previous cesarean section, n (%)	39 (36.1)
History of abortion, n (%)	24 (22.2)
≥2 sexual partners during the previous 12 months, n (%)	13 (12.0)
Sexual intercourse frequency, times/week, median (IQR)	2 (1–3)
Vaginal douching (occasionally or regularly), n (%)	39 (36.1)
Contraceptive method, n (%)	
No contraception	33 (30.6)
Condom	21 (19.4)
Combined oral contraceptive	14 (13.0)
Intrauterine device	26 (24.1)
Withdrawal	11 (10.2)
Other	3 (2.8)
Presenting symptoms§, n (%)	
Vaginal discharge	83 (76.5)
Vulvovaginal itching	68 (62.7)
Malodor	55 (51.0)
Burning sensation	29 (26.9)
Dyspareunia	18 (16.7)
Dysuria	14 (13.0)

*Data are presented as mean±SD, median (IQR), or number (percentage), as appropriate. *Higher education=high school or university. †Middle/high socioeconomic status=middle or high income level. ‡Alcohol use=occasional or regular alcohol consumption. §Participants could report more than one symptom. Values for symptoms are presented as frequencies and percentages; percentages may not sum to 100% because participants could report more than one symptom. SD: standard deviation; IQR: interquartile range.

the preceding 3 months. The most common presenting complaints were vaginal discharge (76.5%), itching (62.7%), and malodor (51.0%).

Based on clinical and laboratory evaluation, 45.4% (n=49) of participants were diagnosed with BV, 29.6% (n=32) with vulvovaginal candidiasis, 10.2% (n=11) with *T. vaginalis* infection, and 11.1% (n=12) with mixed vaginitis. The remaining 3.7% (n=4) were classified as non-specific vaginitis. The four women with non-specific vaginitis were reported descriptively only and were not included in the comparative analyses of vaginitis subtypes.

The median PHS for the overall study population was 7 (interquartile range [IQR]: 6–10), with scores ranging from 3 to 14. Median PHS scores were 6 (IQR: 6–8) in women with BV, 6 (IQR: 5.5–7) in those with trichomoniasis, and 10 (IQR: 8.75–11) in women with vulvovaginal candidiasis. Comparison of PHS scores across diagnostic groups demonstrated lower PHS values in BV and trichomoniasis compared with candidiasis (Kruskal-Wallis test, p<0.001). Figure 1 shows the distribution of PHS scores according to vaginitis subtype.

In the analysis stratified by PHS tertiles, the prevalence of BV (55.1%) and trichomoniasis (54.5%) was higher in the lowest PHS tertile compared with the middle and highest tertiles (p<0.001). In contrast, vulvovaginal candidiasis was more frequently observed in the highest PHS tertile. The distribution of PHS scores and PHS tertiles according to vaginitis subtype is summarized in Table 2.

In univariable logistic regression analysis, each one-point increase in the PHS score was associated with a lower likelihood of BV (OR 0.79, 95%CI 0.67–0.94, p=0.007). Current smoking (OR 3.01, 95%CI 1.26–7.17, p=0.013) and vaginal douching (OR 3.35, 95%CI 1.47–7.61, p=0.004) were also significantly associated with BV in the univariable analysis. The results of the univariable and multivariable logistic regression analyses are presented in Table 3. In the multivariable logistic regression analysis, higher PHS scores remained independently associated with lower odds of BV after adjustment for age, body mass index, smoking status, number of sexual partners during the previous 12 months, and vaginal douching (adjusted OR 0.80, 95%CI 0.67–0.96, p=0.019). Current smoking (adjusted OR 2.90, 95%CI 1.12–7.51, p=0.028) and vaginal douching (adjusted OR 2.91, 95%CI 1.20–7.03, p=0.018) were also independently associated with BV. Body mass index, age, and having two or more sexual partners during the previous 12 months were not independently associated with BV.

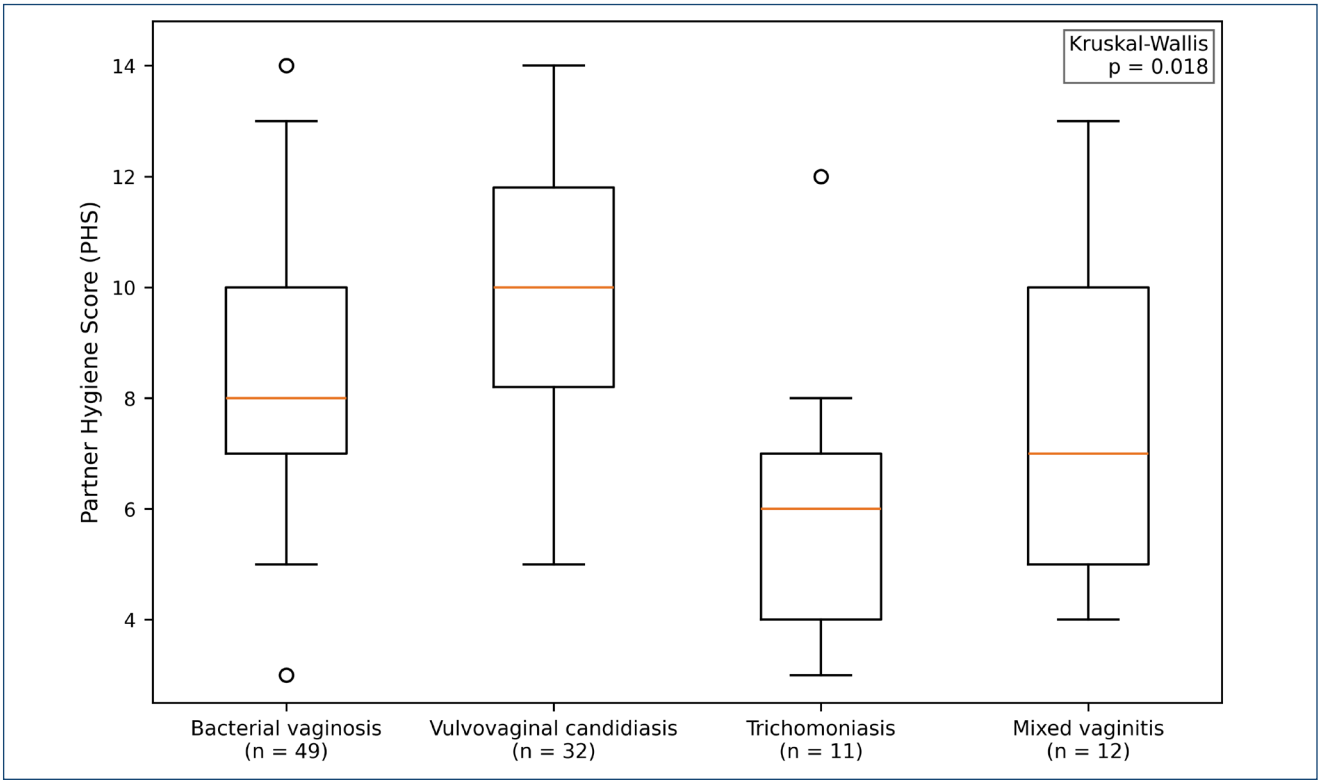


Figure 1. Distribution of partner hygiene score scores across vaginitis subtypes. Boxes represent the interquartile range, the horizontal line indicates the median, whiskers represent values within 1.5 × interquartile range, and circles denote outlier observations. Sample sizes are shown on the x-axis. Differences among groups were assessed using the Kruskal-Wallis test (p<0.001).

Table 2. Partner hygiene score scores and tertile distribution according to vaginitis subtype.

Vaginitis subtype	n	PHS median (IQR)*	PHS range	Low PHS tertile n (%)	Middle PHS tertile n (%)	High PHS tertile n (%)
Bacterial vaginosis	49	6 (6–8)	3–14	27 (55.1)	14 (28.6)	8 (16.3)
Vulvovaginal candidiasis	32	10 (8.75–11)	5–14	6 (18.8)	9 (28.1)	17 (53.1)
Trichomonas vaginalis	11	6 (5.5–7)	3–10	6 (54.5)	4 (36.4)	1 (9.1)
Mixed vaginitis	12	8.5 (6.75–10.25)	4–13	(25.0)	5 (41.7)	4 (33.3)

*Statistical analysis: PHS scores: Kruskal-Wallis test (χ^2=19.89, p<0.001). PHS tertile distribution: Fisher-Freeman-Halton exact test, p=0.003. Fisher-Freeman-Halton exact test was used because some expected cell counts were <5. PHS: partner hygiene score; IQR: interquartile range. PHS tertile cut-points were defined as follows: Low tertile, 3–6; Middle tertile, 7–9; High tertile, 10–14.

Table 3. Univariable and multivariable logistic regression analyses for factors associated with bacterial vaginosis.

Variable	Univariable OR (95%CI)	p	Adjusted OR (95%CI)	p-value
PHS score (per 1-point increase)	**0.79 (0.67–0.94)**	**0.007**	**0.80 (0.67–0.96)**	**0.019**
Age (per year)	0.97 (0.91–1.03)	0.340	0.95 (0.88–1.03)	0.199
Body mass index (per kg/m^2)	0.95 (0.87–1.04)	0.282	0.98 (0.88–1.09)	0.751
Current smoking	**3.01 (1.26–7.17)**	**0.013**	**2.90 (1.12–7.51)**	**0.028**
≥2 sexual partners (previous 12 months)	1.47 (0.46–4.71)	0.515	1.02 (0.27–3.87)	0.971
Vaginal douching (yes)	**3.35 (1.47–7.61)**	**0.004**	**2.91 (1.20–7.03)**	**0.018**

OR: odds ratio; CI: confidence interval; PHS: partner hygiene score. Bold values indicate statistically significant associations (p<0.05); the corresponding ORs and 95%CIs are also bolded for emphasis.

When PHS components were evaluated individually, non-daily partner underwear changing was more frequently reported among women with BV and demonstrated a borderline statistical association (p≈0.08). No statistically significant associations were observed between other PHS components and vaginitis types.

During the 3-month follow-up period, vaginitis recurrence was observed in 33% of cases (n=34). The recurrence rate was higher among women initially diagnosed with BV, followed by mixed vaginitis and *T. vaginalis* infection, while it was lower among those diagnosed with vulvovaginal candidiasis. No statistically significant association was found between overall PHS scores and vaginitis recurrence during the 3-month follow-up period.

DISCUSSION

In this exploratory clinical study, lower partner hygiene—as quantified by a newly developed PHS—was associated with a dysbiosis-predominant vaginitis phenotype. Specifically, PHS scores were significantly lower in women diagnosed with BV and *T. vaginalis* infection compared with VVC, and lower PHS tertiles were enriched for BV and trichomoniasis. In regression analyses, each one-point decrease in PHS was associated with higher odds of BV after adjustment for potential confounders. By contrast, PHS was not associated with 3-month recurrence, despite an overall recurrence rate of approximately one-third.

Although the association between PHS and BV remained statistically significant after multivariable adjustment (adjusted OR 0.80, 95%CI 0.67–0.96), the observed effect size was modest. Therefore, these findings should be interpreted cautiously and require confirmation in larger prospective studies before any clinical implications can be drawn. Although vaginal douching and smoking showed similar effect estimates in the regression model, these factors may partially overlap with broader hygiene-related behaviors and lifestyle patterns. However, vaginal douching and smoking were included as covariates in the multivariable model to minimize potential confounding of the association between PHS and BV.

The recurrence analysis should also be interpreted cautiously, as only 34 recurrence events were observed during follow-up. Consequently, the study may have been underpowered to detect modest associations between baseline PHS and recurrence, and the absence of a statistically significant association should not be interpreted as evidence of no association. Larger prospective studies are warranted to further investigate this relationship.

Our findings support the concept that BV is linked to dysbiosis and partner-associated microbial exchange rather than exclusively host-related susceptibility[5,12,18]. Recent studies using culture-independent molecular techniques have substantially

advanced our understanding of the relationship between the vaginal and penile microbiomes. Using 16S rRNA sequencing, Mehta et al.[6] demonstrated that the composition of the penile microbiome predicts incident BV in female partners. Consistent with these findings, concurrent treatment of male partners has been shown to alter the genital microbiota of couples and may improve treatment outcomes, highlighting the importance of partner-associated microbial reservoirs[19]. Furthermore, prospective cohort studies have shown that sexual activity contributes to the persistence of a non-optimal vaginal microbiota following BV treatment[20], while whole metagenome sequencing has provided direct evidence of microbial exchange between sexual partners[15]. These findings provide a biologically plausible explanation for the observed association between poor partner hygiene and BV, suggesting that poor partner hygiene may contribute to repeated microbial transfer and disruption of the vaginal microbiota. The hypothesized biological pathway linking these mechanisms is summarized in Figure 2. This conceptual model is based on current evidence and is intended to illustrate biological plausibility rather than establish causal relationships.

The particularly low PHS scores observed among women with trichomoniasis may be explained by the sexually transmitted nature of this infection and its established links with vaginal dysbiosis[21-23]. In contrast, VVC is more strongly associated with host-related and metabolic factors influencing the vaginal environment and microbiome composition than partner-mediated bacterial exchange[16,24]. The higher PHS scores among women with VVC in our cohort may therefore reflect a genuine subtype distinction: PHS may be more sensitive to behaviors that facilitate anaerobe-associated dysbiosis and sexually transmitted pathogens than to fungal overgrowth.

Among individual PHS components, non-daily partner underwear changing showed a borderline association with BV. Although not definitive, this is consistent with the hypothesis that specific, easily modifiable behaviors may be associated with microbial persistence or re-seeding between partners. The broader literature suggests that women frequently attribute

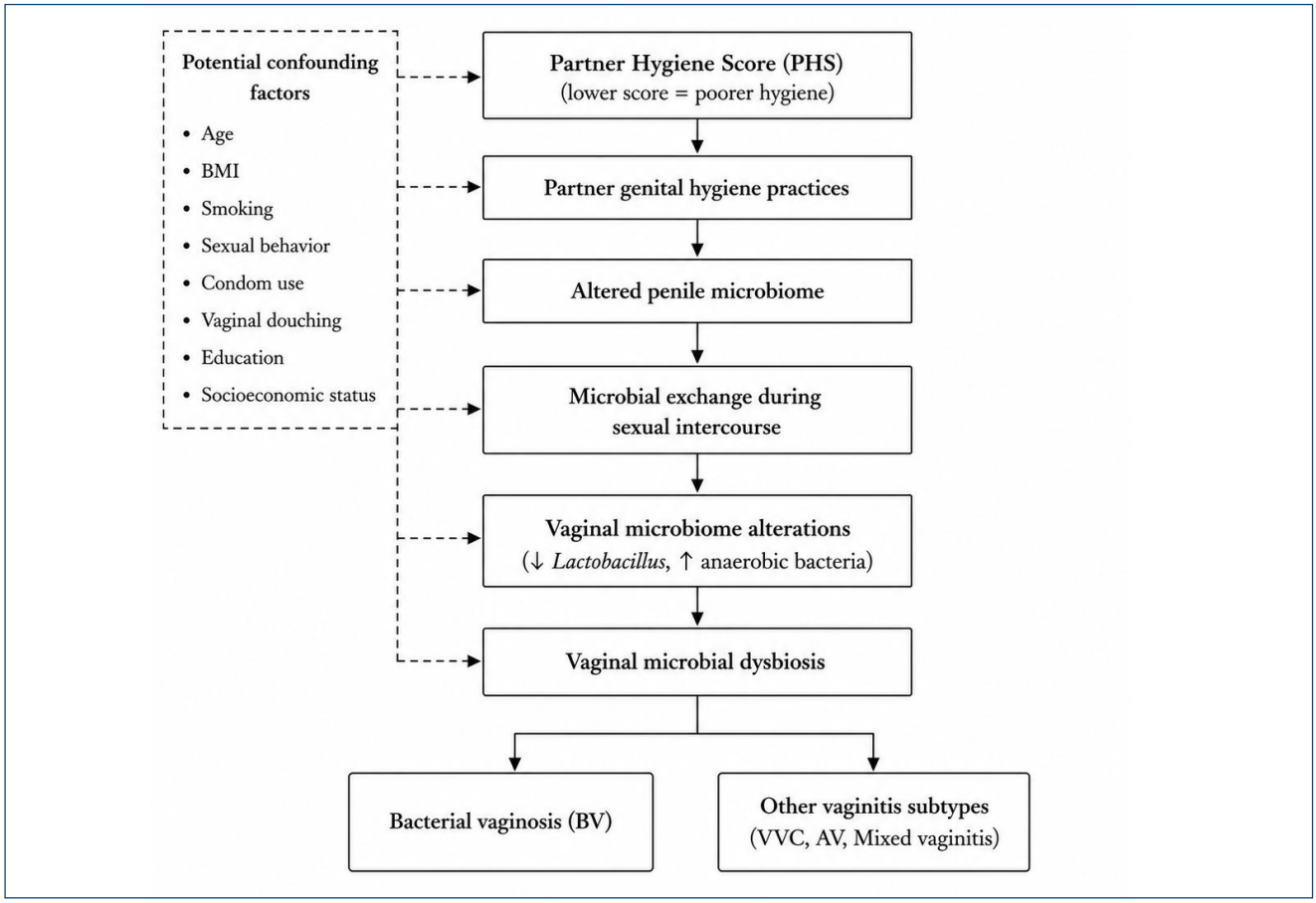


Figure 2. Proposed conceptual model linking partner hygiene to vaginitis subtypes. Solid arrows indicate the hypothesized biological pathway based on current evidence, whereas dashed arrows represent potential confounding factors considered in the statistical analyses. The model is intended to illustrate biological plausibility rather than establish causal relationships.

BV onset and recurrence to sexual behaviors and partner factors, underscoring the clinical relevance of actionable, couple-focused counseling[25]. However, most prior work has evaluated single partner variables (e.g., circumcision status, smoking) rather than aggregating daily hygiene practices into a quantitative measure, limiting comparability across studies[12,13]. Our PHS framework addresses this gap by operationalizing partner hygiene as a multidimensional exposure, enabling modeling of dose–response patterns and subgroup analyses.

Despite evidence that BV is characterized by high recurrence and chronicity in many populations, we did not observe a significant relationship between baseline PHS and recurrence at 3 months[14,18,26]. The lack of association between baseline PHS and recurrence may reflect multifactorial recurrence mechanisms, treatment-related factors, and unmeasured behavioral changes during follow-up[5,9,12,27].

Notably, partner-directed strategies for BV have shown renewed promise, including concurrent male partner treatment and combined antimicrobial approaches that alter couples' genital microbiota[28], although the broader evidence base remains mixed and continues to evolve[27]. Our null recurrence finding should therefore be interpreted cautiously and viewed as hypothesis-refining rather than hypothesis-negating. Beyond partner-directed strategies, emerging evidence suggests that oral probiotics may represent a complementary, partner-independent approach to supporting vaginal health and modulating the vaginal microbiome, particularly in specific populations such as postmenopausal women[24]. However, the role of probiotics in preventing recurrence in premenopausal women with BV remains to be established and warrants investigation alongside partner hygiene interventions.

Clinically, these findings support a more partner-inclusive approach to vaginitis phenotyping—particularly for BV and trichomoniasis—by recognizing that partner behaviors may be associated with exposure to dysbiosis-associated bacterial communities and sexually transmitted pathogens[15,18,23]. While routine "partner hygiene scoring" is premature, the PHS concept could be used to structure targeted counseling on feasible behaviors (e.g., regular underwear changing and consistent genital hygiene around intercourse), alongside evidence-based STI management and prevention.

Although the theoretical PHS range was 0–14, the observed scores in our study ranged from 3 to 14. The absence of scores at the extreme lower end may indicate limited representation of very poor partner hygiene behaviors in this sample; however, formal assessment of floor or ceiling effects was not performed. For research, PHS requires formal validation (reliability, construct validity, and responsiveness) and should be tested in larger, multi-center cohorts with repeated measurement across time. Test–retest reliability of the PHS should be evaluated in future validation studies. Integration of penile and vaginal microbiome profiling would be particularly valuable to determine whether PHS correlates with anaerobe-enriched penile communities or with strain-sharing within couples, strengthening causal inference[6,28]. Finally, intervention studies could examine whether improving PHS-targeted behaviors—alone or combined with partner treatment—are associated with lower BV incidence or recurrence and whether effects differ by baseline subtype and biofilm burden[5,19].

Key strengths include the introduction of a novel, quantitative partner hygiene measure and the subtype-stratified analysis across common vaginitis diagnoses. Limitations include reliance on participant-reported partner behaviors (potential misclassification and social desirability bias), the single-center design, which may limit the generalizability of the findings to other populations and healthcare settings, residual confounding by female intravaginal practices and sexual network factors, and the relatively small sample size, particularly in the trichomoniasis and other smaller diagnostic subgroups, which may have limited the statistical power and precision of subgroup analyses. In addition, partner treatment status, changes in partner hygiene during follow-up, exact time to recurrence, and other time-varying factors potentially associated with recurrence were not prospectively recorded. Therefore, time-to-event analyses such as Cox proportional hazards regression could not be performed, and the recurrence analysis should be interpreted as exploratory. Additionally, the PHS is a newly developed exploratory instrument and has not undergone formal psychometric validation; therefore, its reliability and validity should be confirmed in future studies. Moreover, objective validation of the PHS against external measures, such as partner interviews or microbiological assessments, was not performed in this study. Since p-values were not adjusted for multiple comparisons, significant findings should be interpreted as exploratory.

CONCLUSION

Overall, poorer partner hygiene as captured by PHS was associated with BV and trichomoniasis compared with VVC. These findings indicate subtype-specific cross-sectional associations between partner hygiene and vaginitis phenotypes. The absence of a clear association with 3-month recurrence indicates that recurrence is likely driven by additional factors beyond baseline partner hygiene alone. Given the exploratory nature of the recurrence analysis, these findings should be considered hypothesis-generating and require prospective validation in larger cohorts. Furthermore, before clinical implementation, the PHS requires prospective

validation and formal psychometric evaluation to establish its reliability, validity, and clinical applicability.

ETHICS STATEMENT

The study received approval from the Istanbul Esenyurt University Ethics Committee (numbered [2025/10-21]). All participants provided informed consent prior to enrollment.

AUTHORS' CONTRIBUTIONS

DUK: Conceptualization, Data curation, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. **MEP:** Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – review & editing. UA: Formal Analysis, Methodology, Software, Validation, Writing – review & editing. **MO:** Formal Analysis, Methodology, Resources, Supervision, Writing – review & editing. **HK:** 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. Peebles K, Velloza J, Balkus JE, McClelland RS, Barnabas RV. High global burden and costs of bacterial vaginosis: a systematic review and meta-analysis. Sex Transm Dis. 2019;46(5):304-11. https://doi.org/10.1097/OLQ.0000000000000972
2. Rafat Z, Alipour N, Esmaeili E, Mirhasani F, Yaseri M, Haghighat F, et al. Vulvovaginal candidiasis: a global perspective through systematic review and meta-analysis. Microb Pathog. 2026;212:108284. https://doi.org/10.1016/j.micpath.2026.108284
3. Rowley J, Vander Hoorn S, Korenromp E, Low N, Unemo M, Abu-Raddad LJ, et al. Chlamydia, gonorrhoea, trichomoniasis and syphilis: global prevalence and incidence estimates, 2016. Bull World Health Organ. 2019;97(8):548-62P. https://doi.org/10.2471/BLT.18.228486
4. Reiter S, Kellogg Spadt S. Bacterial vaginosis: a primer for clinicians. Postgrad Med. 2019;131(1):8-18. https://doi.org/10.1080/00325481.2019.1546534
5. Swidsinski A, Amann R, Guschin A, Swidsinski S, Loening-Baucke V, Mendling W, et al. Polymicrobial consortia in the pathogenesis of biofilm vaginosis visualized by FISH. Historic review outlining the basic principles of the polymicrobial infection theory. Microbes Infect. 2024;26(8):105403. https://doi.org/10.1016/j.micinf.2024.105403
6. Mehta SD, Agingu W, Nordgren RK, Green SJ, Bhaumik DK, Bailey RC, et al. Characteristics of women and their male sex partners predict bacterial vaginosis among a prospective cohort of kenyan women with nonoptimal vaginal microbiota. Sex Transm Dis. 2020;47(12):840-50. https://doi.org/10.1097/OLQ.0000000000001259
7. Price LB, Liu CM, Johnson KE, Aziz M, Lau MK, Bowers J, et al. The effects of circumcision on the penis microbiome. PLoS One. 2010;5(1):e8422. https://doi.org/10.1371/journal.pone.0008422
8. Zenilman JM, Fresia A, Berger B, McCormack WM. Bacterial vaginosis is not associated with circumcision status of the current male partner. Sex Transm Infect. 1999;75(5):347-8. https://doi.org/10.1136/sti.75.5.347
9. Alcaide ML, Cook R, Chisembele M, Malupande E, Jones DL. Determinants of intravaginal practices among HIV-infected women in Zambia using conjoint analysis. Int J STD AIDS. 2016;27(6):453-61. https://doi.org/10.1177/0956462415585447
10. Rodriguez VJ, Chisembele M, Jones DL, Cook R, Weiss SM, Alcaide ML. Influencing the importance of health, partners, and hygiene among Zambian women. Int J STD AIDS. 2018;29(3):259-65. https://doi.org/10.1177/0956462417723546
11. Mehta SD, Nordgren RK, Agingu W, Otieno F, Odongo W, Odhiambo F, et al. Sexual quality of life and association with HIV and sexually transmitted infections among a cohort of heterosexual couples in Kenya. J Sex Med. 2018;15(10):1446-55. https://doi.org/10.1016/j.jsxm.2018.08.007
12. Verstraelen H, Verhelst R, Vaneechoutte M, Temmerman M. The epidemiology of bacterial vaginosis in relation to sexual behaviour. BMC Infect Dis. 2010;10:81. https://doi.org/10.1186/1471-2334-10-81
13. Bukusi EA, Cohen CR, Meier AS, Waiyaki PG, Nguti R, Njeri JN, et al. Bacterial vaginosis: risk factors among Kenyan women and their male partners. Sex Transm Dis. 2006;33(6):361-7. https://doi.org/10.1097/01.olq.0000200551.07573.df
14. Francis SC, Looker C, Vandepitte J, Bukenya J, Mayanja Y, Nakubulwa S, et al. Bacterial vaginosis among women at high risk for HIV in Uganda: high rate of recurrent diagnosis despite treatment. Sex Transm Infect. 2016;92(2):142-8. https://doi.org/10.1136/sextrans-2015-052160
15. Carter KA, France MT, Rutt L, Bilski L, Martinez-Greiwe S, Regan M, et al. Sexual transmission of urogenital bacteria: whole metagenome sequencing evidence from a sexual network study. mSphere. 2024;9(3):e0003024. https://doi.org/10.1128/msphere.00030-24
16. Zeng X, Zhang Y, Zhang T, Xue Y, Xu H, An R. Risk factors of vulvovaginal candidiasis among women of reproductive age in Xi'an: a cross-sectional study. Biomed Res Int. 2018;2018:9703754. https://doi.org/10.1155/2018/9703754
17. Biswas R, Thoma M, Kong X. Functional data analysis to characterize disease patterns in frequent longitudinal data: application to bacterial vaginal microbiota patterns using weekly Nugent scores and identification of pattern-specific risk factors. BMC Med Res Methodol. 2023;23(1):251. https://doi.org/10.1186/s12874-023-02063-8

18. Bradshaw CS, Plummer EL, Muzny CA, Mitchell CM, Fredricks DN, Herbst-Kralovetz MM, et al. Bacterial vaginosis. Nat Rev Dis Primers. 2025;11(1):43. https://doi.org/10.1038/s41572-025-00626-1
19. Plummer EL, Vodstrcil LA, Doyle M, Danielewski JA, Murray GL, Fehler G, et al. A prospective, open-label pilot study of concurrent male partner treatment for bacterial vaginosis. mBio. 2021;12(5):e0232321. https://doi.org/10.1128/mBio.02323-21
20. Ratten LK, Plummer EL, Murray GL, Danielewski J, Fairley CK, Garland SM, et al. Sex is associated with the persistence of non-optimal vaginal microbiota following treatment for bacterial vaginosis: a prospective cohort study. BJOG. 2021;128(4):756-67. https://doi.org/10.1111/1471-0528.16430
21. Bochner AF, Baeten JM, Rustagi AS, Nakku-Joloba E, Lingappa JR, Mugo NR, et al. A cross-sectional analysis of *Trichomonas vaginalis* infection among heterosexual HIV-1 serodiscordant African couples. Sex Transm Infect. 2017;93(7):520-9. https://doi.org/10.1136/sextrans-2016-053034
22. Jewanraj J, Ngcapu S, Liebenberg LJP. Semen: a modulator of female genital tract inflammation and a vector for HIV-1 transmission. Am J Reprod Immunol. 2021;86(5):e13478. https://doi.org/10.1111/aji.13478
23. Waltmann A, Thomas C, Duncan JA. The role of the genital microbiota in the acquisition and pathogenesis of sexually transmitted infections. Curr Opin Infect Dis. 2023;36(1):35-48. https://doi.org/10.1097/QCO.0000000000000893
24. Silva VF, Refinetti P, Vicariotto F, Baracat EC, Soares Junior JM. Oral probiotics and vaginal microbiome in post-menopause women: an opinion for the improvement of natural therapies in gynecology. Rev Assoc Med Bras (1992). 2024;70(2):e702EDIT. https://doi.org/10.1590/1806-9282.702EDIT
25. Bilardi JE, Walker SM, Temple-Smith MJ, McNair RP, Mooney-Somers J, Vodstrcil LA, et al. Women view key sexual behaviours as the trigger for the onset and recurrence of bacterial vaginosis. PLoS One. 2017;12(3):e0173637. https://doi.org/10.1371/journal.pone.0173637
26. Marshall AO. Managing recurrent bacterial vaginosis: insights for busy providers. Sex Med Rev. 2015;3(2):88-92. https://doi.org/10.1002/smrj.45
27. Mehta SD. Systematic review of randomized trials of treatment of male sexual partners for improved bacteria vaginosis outcomes in women. Sex Transm Dis. 2012;39(10):822-30. https://doi.org/10.1097/OLQ.0b013e3182631d89
28. Zozaya M, Ferris MJ, Siren JD, Lillis R, Myers L, Nsuami MJ, et al. Bacterial communities in penile skin, male urethra, and vaginas of heterosexual couples with and without bacterial vaginosis. Microbiome. 2016;4:16. https://doi.org/10.1186/s40168-016-0161-6

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260869

Comparative evaluation of ChatGPT and Gemini feedback in scenario-based clinical reasoning among medical interns

Hilal Hatice Ülkü[1]*, Fulya Torun[2], Selcen Öncü[3]

SUMMARY

OBJECTIVE: Artificial intelligence tools are increasingly used to provide individualized feedback in medical education. The aim of this study was to compare the quality of feedback generated by ChatGPT and Gemini during scenario-based clinical reasoning activities.

METHODS: A convergent mixed-methods study was conducted with 32 medical interns randomly assigned to ChatGPT (n=16) or Gemini (n=16). Following a virtual patient interview on iron deficiency anemia, artificial intelligence-generated feedback was evaluated by interns and medical education specialists using an 11-item rubric assessing clarity, accuracy, structure, identification of strengths and weaknesses, contribution to clinical reasoning, and overall educational value. Quantitative data were analyzed using descriptive statistics and Mann-Whitney U tests, while qualitative data were examined through content analysis.

RESULTS: Interns rated ChatGPT slightly higher than Gemini (1.80±0.20 vs. 1.75±0.26), whereas specialists reported similar scores (1.91±0.06 vs. 1.87±0.10). No statistically significant differences were identified (p>0.05). Both systems were perceived as clear, supportive, and constructive; however, insufficient patient-centeredness emerged as a common limitation of both artificial intelligence systems. Qualitative findings indicated that ChatGPT provided more structured and reflective feedback, whereas Gemini offered more concrete examples. Common limitations included insufficient patient-centeredness, lack of systematicity, and occasional unrealistic statements.

CONCLUSION: ChatGPT and Gemini generated educationally useful feedback for scenario-based clinical reasoning activities. Although artificial intelligence tools cannot replace expert supervision, they may support personalized learning and reduce faculty workload when used under human oversight.

KEYWORDS: Artificial intelligence. Clinical reasoning. Feedback. Medical education. Patient simulation.

BACKGROUND

Clinical reasoning is a core competency in medical education and is particularly important during internship, when students integrate theoretical knowledge with patient care responsibilities. Scenario-based case analysis is widely used to support clinical reasoning and decision-making skills, and effective feedback is essential for maximizing learning outcomes. High-quality feedback helps learners identify strengths and weaknesses, promotes reflection, and supports professional development[1-3].

Recent advances in artificial intelligence (AI), particularly large language models (LLMs), have created new opportunities for providing individualized educational feedback. AI-based systems can analyze learner responses, generate personalized comments, and offer suggestions for improvement in a timely manner. These features may be especially valuable in medical education, where increasing student numbers and faculty workload often limit opportunities for individualized feedback[4-7].

Despite these advantages, concerns remain regarding the educational quality of AI-generated feedback. Previous studies have highlighted issues related to contextual understanding, consistency, transparency, and the ability of AI systems to accurately reflect clinical performance. Therefore, AI-generated feedback should be evaluated not only from a technical perspective but also in terms of its educational value and relevance to clinical learning[8-10].

Among the most widely used AI tools, ChatGPT and Gemini have demonstrated considerable potential in supporting learning activities. However, despite their widespread use, evidence directly comparing the educational quality of feedback generated by these two LLMs during scenario-based clinical reasoning activities remains scarce. To the best of our knowledge, no previous study has directly compared ChatGPT and Gemini in this specific educational context from the perspectives of both medical interns and medical education specialists. Understanding how learners and educators perceive

[1]Aydın Adnan Menderes University, Aydın Vocational School – Aydın, Turkey.
[2]Aydın Adnan Menderes University, Faculty of Education, Department of Computer and Instructional Technologies – Aydın, Turkey.
[3]Aydın Adnan Menderes University, Faculty of Medicine, Department of Medical Education – Aydın, Turkey.
*Corresponding author: hilal.gulludere@adu.edu.tr
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on June 26, 2026. Accepted on August 01, 2026.
Scientific Editor: José Maria Soares Júnior

AI-generated feedback is important for determining the role of these technologies in medical education.

Therefore, this study aimed to compare the quality of feedback generated by ChatGPT and Gemini during scenario-based clinical reasoning activities among medical interns. Specifically, feedback quality was evaluated by both medical interns and medical education specialists, and qualitative analyses were conducted to identify strengths and areas for improvement.

METHODS

Study design

A convergent mixed-methods study was conducted to evaluate the quality of feedback generated by two AI tools, ChatGPT and Gemini, during scenario-based clinical reasoning activities. Quantitative and qualitative data were collected concurrently and integrated during interpretation.

Participants

The study included 32 final-year medical interns from Aydın Adnan Menderes University, Türkiye. Participants were randomly assigned to either ChatGPT (n=16) or Gemini (n=16). Because this was an exploratory educational study, a convenience sample consisting of all eligible interns who volunteered during the study period was included. No formal a priori sample size calculation or power analysis was performed. In addition, two medical education specialists independently evaluated all AI-generated feedback.

Data collection

A structured virtual patient scenario on iron deficiency anemia was developed based on national undergraduate medical education competencies and reviewed by experts in internal medicine and family medicine. A standardized virtual patient scenario, interview prompt, and feedback prompt were developed and pilot-tested before implementation to ensure consistency across AI interactions. The complete scenario and standardized prompts used for both ChatGPT-4o and Gemini 2.5 Flash are provided in Supplementary File 1.

ChatGPT-4o and Gemini 2.5 Flash were used as AI tools. Each participant completed a written virtual patient interview with the assigned AI system. Following the interaction, the AI generated individualized feedback regarding the participant's clinical reasoning and patient interview performance. These AI-generated feedback reports subsequently constituted the evaluation material for both the interns and the medical education specialists.

Feedback quality was assessed using a researcher-developed rubric that was developed based on established feedback quality frameworks in medical education and reviewed by two medical education experts for content validity. The rubric consisted of 11 items evaluating clarity, accuracy, structure, identification of strengths and weaknesses, contribution to clinical reasoning, and overall educational value. Items were scored as "No" (0), "Partly" (1), or "Yes" (2), with higher scores indicating better feedback quality.

The same 11-item rubric was completed by both the medical interns and the medical education specialists. However, their evaluation perspectives differed. Interns evaluated the quality of the AI-generated feedback they personally received following the virtual patient interview, whereas the specialists independently evaluated the educational quality of the AI-generated feedback itself. To minimize potential evaluator bias, all feedback reports were anonymized and presented to the specialists by an independent researcher who was not involved in the evaluation process. Consequently, the specialists remained blinded to whether each feedback report had been generated by ChatGPT-4o or Gemini 2.5 Flash.

For the qualitative component, evaluators provided written comments regarding the overall quality, strengths, and areas for improvement of the feedback.

Statistical analysis

Descriptive statistics were calculated for all variables. Because data were not normally distributed, comparisons between ChatGPT and Gemini were performed using the Mann-Whitney U test. Statistical significance was set at $p<0.05$. Analyses were conducted using SPSS software.

Qualitative data were analyzed using content analysis with MAXQDA 2022. Two researchers independently coded the data and discussed discrepancies until consensus was achieved. Percent agreement between coders was 98.3%.

RESULTS

In total, 32 medical interns participated in the study (ChatGPT=16, Gemini=16). Feedback quality was evaluated by both interns and medical education specialists.

Interns rated ChatGPT slightly higher than Gemini (1.80±0.20 vs. 1.75±0.26), while specialists reported similar scores (1.91±0.06 vs. 1.87±0.10). No statistically significant differences were observed between the two AI tools in either intern or specialist evaluations (all $p>0.05$) (Table 1). Effect size analysis showed a small difference in intern evaluations (Cohen's $d=0.22$) and a moderate difference favoring ChatGPT in specialist evaluations (Cohen's $d=0.49$).

Among intern evaluations, the highest-rated characteristics of AI-generated feedback were the use of non-judgmental language, supportive communication, and clear identification of strengths. The lowest-rated rubric criterion for both systems was Descriptiveness, reflecting the tendency of the AI-generated feedback to include inferences beyond information explicitly stated during the interview. Among specialist evaluations, the highest-rated rubric criteria were Clarity and Comprehensibility, Descriptiveness, and Justification and Rationale, whereas Scientific Accuracy and Organization (Systematic Structure) received the lowest ratings for both ChatGPT and Gemini (Table 2).

Qualitative analysis identified three major domains: overall feedback quality, strengths, and areas for improvement. Feedback generated by both AI systems was generally perceived as clear, motivating, constructive, and educationally useful. Participants particularly valued the detailed explanations and practical suggestions provided by the systems.

However, several limitations were identified. The most frequently reported concerns included insufficient detail, lack of systematic organization, limited emphasis on communication skills, and inadequate identification of clinical errors. Some feedback was also perceived as unrealistic or insufficiently patient-centered.

Comparison of the two systems revealed distinct characteristics. ChatGPT was perceived as more structured, concise, and reflective, frequently encouraging learners to reconsider their reasoning process. Gemini was viewed as providing more concrete examples and realistic conversational interactions, although patient-centered questioning was less frequently emphasized. A summary of the qualitative findings is presented in Figure 1.

Table 1. Interns and medical education specialists' evaluation of artificial intelligence-generated feedback (mean±standard deviation).

Evaluator	ChatGPT	Gemini	p-value	Cohen's d
Interns	1.80±0.20	1.75±0.26	0.552	0.22
Specialists	1.91±0.06	1.87±0.10	0.410	0.49

Table 2. Items with the highest and lowest scores.

Evaluator	Highest-rated items	Lowest-rated items
Interns	Supportive language Identification of strengths	Descriptiveness (feedback based only on the student's responses without unsupported inferences)
Specialists	Clarity and comprehensibility; contextual appropriateness (descriptiveness); justification and rationale	Scientific accuracy; organization (systematic structure)

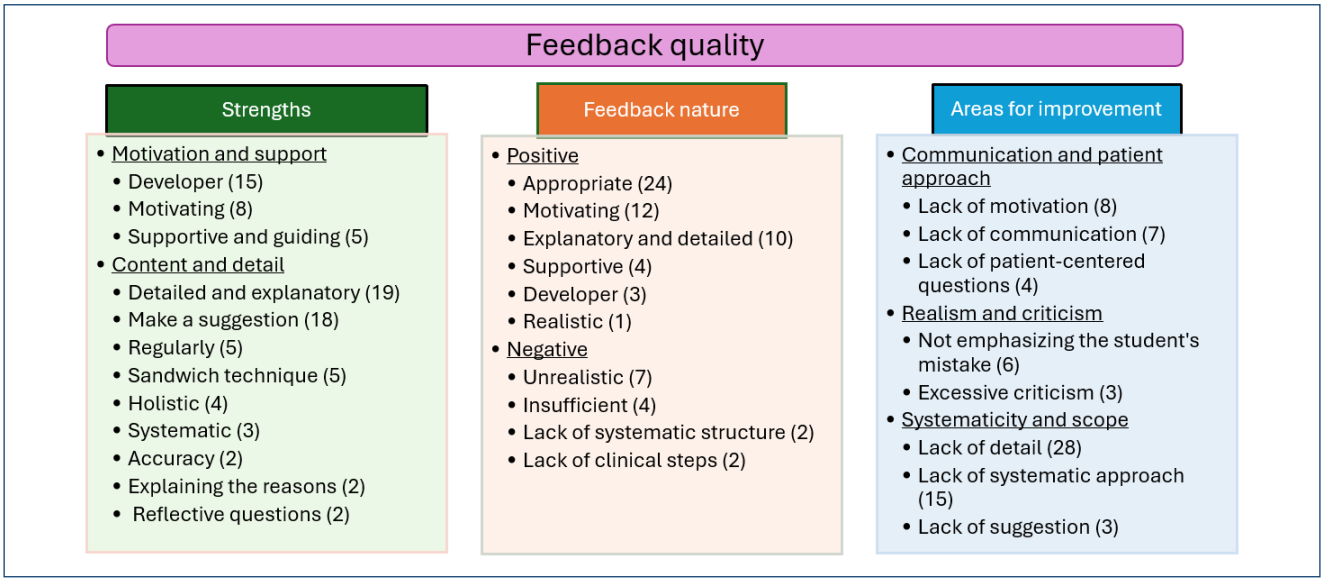


Figure 1. Code distribution of feedback quality.

DISCUSSION

This study evaluated the quality of feedback generated by ChatGPT and Gemini during scenario-based clinical reasoning activities. Both interns and medical education specialists perceived the feedback as clear, supportive, constructive, and educationally valuable. Although no statistically significant differences were identified between the two AI systems, integrating the quantitative and qualitative findings revealed important differences in the educational characteristics of the feedback generated by each system. While the overall rubric scores were comparable, the qualitative findings provided additional insight into the specific strengths and limitations underlying these quantitative results.

One of the most important findings was that both AI tools were rated highly for clarity, supportive language, and identification of strengths and areas for improvement. Feedback is a critical component of clinical education because it enables learners to recognize performance gaps, reflect on their decisions, and improve future performance. High-quality feedback has been associated with improved self-regulation, motivation, and professional development, whereas poorly structured feedback may reduce learner engagement and limit educational benefit. The positive evaluations obtained from both interns and specialists suggest that AI systems can provide educationally meaningful feedback that aligns with key principles of effective feedback in medical education[3].

Another notable finding was the perceived personalization of AI-generated feedback. Participants reported that feedback was generally relevant to their performance and frequently included suggestions for improvement. This finding is consistent with previous studies demonstrating that AI systems can adapt feedback to learner responses and support individualized learning experiences[11,12]. Personalized feedback is particularly valuable in medical education, where students progress at different rates and require targeted guidance to develop clinical reasoning skills. The ability of AI systems to provide immediate and individualized responses may help address challenges associated with large student cohorts and limited faculty availability.

The qualitative findings further highlighted the educational value of AI-generated feedback. Both systems were frequently described as explanatory, motivating, and developmental. Participants appreciated feedback that not only identified deficiencies but also offered practical recommendations for improvement. Similar findings have been reported in studies showing that ChatGPT-generated feedback may be perceived as more detailed and easier to understand than traditional feedback approaches[12,13]. These characteristics suggest that AI-based systems may complement existing educational practices by promoting reflection and supporting self-directed learning.

Despite these strengths, several limitations of AI-generated feedback were identified. The lowest-rated item among interns concerned the tendency of AI systems to generate interpretations that extended beyond information explicitly provided during the consultation. This finding suggests that AI systems may occasionally produce assumptions or inferences that are not fully supported by the available data. Similar concerns have been reported in previous studies evaluating LLMs in educational settings[14]. From an educational perspective, such inaccuracies may reduce learner trust and potentially reinforce misconceptions if feedback is accepted uncritically. More importantly, feedback based on assumptions rather than the learner's actual performance may lead to inaccurate educational judgments and inappropriate learning recommendations. These findings reinforce the need for human oversight when AI-generated feedback is used in medical education, particularly in contexts involving formative assessment and clinical reasoning. Therefore, AI-generated feedback should be interpreted carefully and verified before being incorporated into educational decision-making or learner feedback.

Participants also identified limitations related to systematic organization, depth of feedback, and patient-centeredness. Effective clinical feedback should address not only diagnostic reasoning but also communication skills, patient-centered care, and professional behavior. In the present study, both systems were perceived as less effective in emphasizing patient-centered questioning and communication-related aspects of clinical encounters. Given the increasing emphasis on patient-centered care in contemporary medical education, this limitation deserves particular attention. Since communication and patient-centered care are core competencies in competency-based medical education, limitations in these areas may reduce the educational usefulness of AI-generated feedback. AI-generated feedback may be more effective when supplemented by educator guidance that explicitly addresses communication and professionalism competencies.

Another important concern was the limited ability of AI systems to identify and emphasize clinical errors. One of the primary purposes of feedback in clinical education is to help learners recognize mistakes before encountering similar situations in real clinical practice. However, qualitative findings indicated that some errors were insufficiently highlighted or inadequately discussed. Similar limitations have been reported in AI-supported problem-based learning environments, where contextual understanding and clinical nuance remain challenging

for LLMs[4]. These findings reinforce the need for human oversight when AI-generated feedback is used in medical training.

Although quantitative comparisons revealed no statistically significant differences between ChatGPT and Gemini, qualitative findings suggested distinct educational characteristics. ChatGPT was generally perceived as more structured, reflective, and supportive. In particular, its tendency to encourage reflection may be beneficial for the development of clinical reasoning and self-assessment skills. Gemini, in contrast, was perceived as providing more concrete examples and realistic conversational interactions. However, patient-centered questioning was less frequently emphasized. These findings suggest that different AI systems may possess unique educational strengths and that their suitability may vary according to instructional goals. Although quantitative comparisons revealed no statistically significant differences, effect size estimates suggested a moderate advantage for ChatGPT in specialist evaluations, indicating potentially greater educational value that warrants further investigation. These qualitative findings also help explain the quantitative results. Although overall rubric scores did not differ significantly between the two AI systems, specialists identified systematic organization as one of the lowest-rated aspects of AI-generated feedback. This finding aligns with the qualitative perception that ChatGPT produced more structured and reflective feedback, which may explain the moderate effect size favoring ChatGPT in specialist evaluations. Likewise, Gemini's emphasis on providing concrete examples and more natural conversational interactions may have contributed to its comparable overall rubric scores despite differences in specific educational characteristics.

The educational implications of these findings are substantial. Previous research has shown that AI-supported feedback systems can improve learning efficiency, facilitate personalized instruction, and reduce faculty workload[15-17]. Our findings support this growing body of evidence and indicate that AI-generated feedback can serve as a valuable supplementary resource in clinical education. However, the limitations identified in this study demonstrate that AI systems should not replace expert educators. Rather, they should be integrated within a hybrid educational model in which AI-generated feedback is reviewed, contextualized, and enriched by human experts. Such an approach may maximize the benefits of AI while minimizing risks associated with inaccuracies, lack of contextual understanding, and insufficient patient-centeredness. In practice, this hybrid model could be implemented by using AI-generated feedback as an initial formative review immediately following a student's virtual patient interview or clinical case presentation. The AI system could provide rapid, structured feedback by identifying strengths, areas for improvement, and preliminary recommendations, after which a faculty member would review, verify, and expand the feedback to ensure scientific accuracy, educational relevance, patient-centeredness, and alignment with curricular learning objectives. This collaborative workflow preserves the educational value of timely AI-supported feedback while maintaining the essential role of educators in fostering clinical reasoning, reflective practice, and professional development.

LIMITATIONS

This study was conducted at a single medical school with a relatively small convenience sample of interns and a limited number of medical education specialists. In addition, no formal a priori sample size calculation or power analysis was performed. Therefore, the findings should be interpreted with appropriate caution and confirmed in larger, adequately powered studies. In addition, AI-generated outputs are highly dependent on prompt design and model version. Because commercial AI systems evolve rapidly, future versions of ChatGPT and Gemini may produce different results. Another limitation is that this study evaluated the perceived educational quality of AI-generated feedback rather than its direct impact on learning outcomes. Although participants generally perceived the feedback as educationally valuable, the study did not determine whether it resulted in measurable improvements in clinical reasoning, knowledge acquisition, or subsequent learner performance. Furthermore, occasional role-assignment errors and unrealistic responses required researcher intervention during implementation. Future studies should include larger multi-center samples and examine the effects of AI-generated feedback on clinical performance, clinical reasoning, and long-term learning outcomes.

CONCLUSION

Both ChatGPT- and Gemini-generated feedback were perceived as clear, supportive, and educationally useful by medical interns and medical education specialists. No significant differences were identified between the two systems. Nevertheless, qualitative findings suggested that ChatGPT provided more structured and reflective feedback, whereas Gemini offered more concrete examples. Although AI-generated feedback shows promise as a tool for supporting personalized learning and reducing faculty workload, limitations related to contextual accuracy, patient-centeredness, and consistency remain. Therefore, AI-based feedback should be integrated into medical

education not as an independent assessment strategy, but as a complementary resource supported by expert oversight.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the Non-Interventional Ethics Committee of the Faculty of Medicine, Aydın Adnan Menderes University (Prot. No: 2025/142-35). Before data collection, participants were provided with information regarding the voluntary basis of their involvement and the use of their responses solely for research purposes. Informed consent was subsequently obtained, and participants were assured that all data would remain confidential. The study procedures adhered strictly to applicable ethical standards and regulatory requirements.

AUTHORS' CONTRIBUTIONS

HHÜ: Conceptualization, Data curation, Formal Analysis, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **FT:** Conceptualization, Data curation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **SÖ:** Conceptualization, Data curation, Formal analysis, Methodology, 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. Eva KW. What every teacher needs to know about clinical reasoning. Med Educ. 2005;39(1):98-106.
2. Gruppen LD. Clinical reasoning: defining it, teaching it, assessing it, studying it. West J Emerg Med. 2016;18(1):4.
3. Hattie J, Timperley H. The power of feedback. Rev Educ Res. 2007;77(1):81-112.
4. Hui Z, Zewu Z, Jiao H, Yu C. Application of ChatGPT-assisted problem-based learning teaching method in clinical medical education. BMC Med Educ. 2025;25:50.
5. Ingley SJ, Pack A. Leveraging AI tools to develop the writer rather than the writing. Trends Ecol Evol. 2023;38(9):785-7.
6. Chiu TK, Xia Q, Zhou X, Chai CS, Cheng M. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. Comput Educ Artif Intell. 2023;4:100118.
7. Farrokhnia M, Banihashem SK, Noroozi O, Wals A. A SWOT analysis of ChatGPT: implications for educational practice and research. Innov Educ Teach Int. 2024;61(3):460-74.
8. Zhu C, Sun M, Luo J, Li T, Wang M. How to harness the potential of ChatGPT in education? Knowl Manag E-Learn. 2023;15(2):133-52.
9. Wartman SA, Combs CD. Reimagining medical education in the age of AI. AMA J Ethics. 2019;21(2):E146-52.
10. Kung TH, Cheatham M, Medenilla A, Sillos C, Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLoS Digit Health. 2023;2(2):e0000198.
11. Doraiswamy PM, Blease C, Bodner K. Artificial intelligence and the future of psychiatry: insights from a global physician survey. Artif Intell Med. 2020;102:101753.
12. Dai W, Lin J, Jin H, Li T, Tsai YS, Gasevic D, et al. Can large language models provide feedback to students? A case study on ChatGPT. In: Chen NS, Rudolph G, Sampson DG, Chang M, Kuo R, Tlili A, editors. Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT 2023). Piscataway (NJ): IEEE; 2023. p. 323-5.
13. Mizumoto A, Eguchi M. Exploring the potential of using an AI language model for automated essay scoring. Res Methods Appl Linguist. 2023;2(2):100050.
14. Khan SA, Taiyara J, Zary N, Otaki F. Artificial intelligence in narrative feedback analysis for competency-based medical education: a review. Stud Health Technol Inform. 2025;327:1423-7.
15. Kewalramani D, Roman DS, Lagos SA, Smith MA, Stern JR, Patel VI, et al. Quality matters: artificial intelligence-based assessment of feedback quality predicts technical skill improvement. Surgery. 2025;187:109625.
16. Othman N, Matore MEEM, Yunus FW. Artificial intelligence in educational measurement: a bibliometric review (1997 to 2024). J Appl Sci Eng Technol Educ. 2025;7(2):314-25.
17. Alfredo R, Echeverria V, Jin Y, Martinez-Maldonado R, Buckingham Shum S, Gašević D, et al. Human-centred learning analytics and AI in education: a systematic literature review. Comput Educ Artif Intell. 2024;6:100215.

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260876

Subacute granulomatous thyroiditis: histopathological spectrum, perineural inflammation, and coexisting lesions

Muhammed Yunus Akar[1], Hasan Gucer[1]*

SUMMARY

OBJECTIVE: Subacute granulomatous thyroiditis is a self-limited inflammatory disorder and the most common cause of painful thyroid disease. Surgically treated cases offer a unique opportunity to explore its histopathological spectrum and co-occurring thyroid pathologies.

METHODS: We retrospectively analyzed 44 thyroidectomy specimens diagnosed with subacute granulomatous thyroiditis between 2009 and 2019, evaluating clinical data and detailed histopathological features including inflammation extent and associated thyroid lesions.

RESULTS: The cohort comprised 36 females and 8 males (mean age 48.1±8.9 years). Preoperative fine-needle aspiration cytology was performed in 68.2% but correctly suggested subacute granulomatous thyroiditis in only one case. Surgery was most indicated for multinodular goiter (63.6%). All cases exhibited confluent folliculocentric granulomas; additional findings included fibrosis (90.9%), neutrophilic infiltration (63.6%), follicular atrophy (63.6%), and perineural inflammation (63.6%). Coexisting follicular nodular disease was present in 88.6% and chronic lymphocytic thyroiditis in 15.9%. Concomitant malignancy was identified in 13 cases (29.5%), predominantly subcentimeter papillary thyroid carcinomas, with no significant clinicopathological associations ($p>0.05$).

CONCLUSION: Subacute granulomatous thyroiditis demonstrates broader pathological significance than traditionally recognized. Perineural inflammation may contribute to pain generation, and the elevated malignancy rate most likely reflects incidental detection due to extensive sampling. These findings highlight diagnostic pitfalls and underscore the importance of careful pathological assessment.

KEYWORDS: De Quervain thyroiditis. Thyroiditis, subacute. Peripheral nerves. Inflammation. Thyroid diseases. Thyroid cancer, papillary. Papillary carcinoma.

INTRODUCTION

Subacute granulomatous thyroiditis (SGT), also known as de Quervain thyroiditis, is a self-limited post-viral inflammatory disorder of the thyroid gland and the most common cause of painful thyroid disease. It predominantly affects middle-aged women, and most patients are managed conservatively without surgical intervention[1].

Despite well-recognized clinical features, the histopathological spectrum of SGT remains relatively underexplored, as surgical specimens are uncommon and typically limited to diagnostically challenging cases[2]. SGT can clinically, radiologically, and macroscopically mimic thyroid malignancy[3-5], and its coexistence with nodular thyroid disease may obscure the detection of underlying neoplasms.

Although the classical hallmark of SGT is granulomatous inflammation with multinucleated giant cells, additional features such as fibrosis, neutrophilic infiltration, and follicular destruction vary according to disease stage. Notably, two aspects remain incompletely understood: the mechanism underlying its characteristic pain and its relationship with coexisting thyroid malignancies. Inflammatory processes involving peripheral nerves have been implicated in pain generation in various organ systems[6], yet this mechanism has not been systematically evaluated in SGT. Similarly, the increasing detection of incidental subcentimeter papillary thyroid carcinomas raises questions about whether any reported association with SGT reflects a true biological relationship.

This study aimed to provide a comprehensive clinicopathological evaluation of SGT in thyroidectomy specimens, with particular emphasis on (i) detailed histopathological characterization including less commonly described findings; (ii) the presence and potential significance of peripheral nerve inflammation; and (iii) the frequency and nature of coexisting benign and malignant thyroid lesions. We also sought to highlight diagnostic pitfalls of preoperative fine-needle aspiration cytology (FNAC).

To our knowledge, no previous surgical pathology series has systematically evaluated the full histopathological spectrum of

[1]Recep Tayyip Erdogan University, Faculty of Medicine, Department of Pathology – Rize, Türkiye.
*Corresponding author: hasan.gucer@erdogan.edu.tr, hasangucer@gmail.com
Conflicts of interest: the authors declare there are no conflicts of interest. Funding: none.
Received on June 07, 2026. Accepted on July 11, 2026.
Scientific Editor: José Maria Soares Júnior

SGT, including perineural inflammation and associated thyroid lesions. Furthermore, the mechanism underlying thyroid pain and the significance of coexisting thyroid carcinoma remain incompletely understood. These knowledge gaps formed the rationale for the present study.

METHODS

Case selection and study design

A retrospective search of the hospital information management system identified 44 cases diagnosed with SGT among total thyroidectomy specimens received at the Medical Pathology Laboratory of Recep Tayyip Erdoğan University Training and Research Hospital between 2009 and 2019. Ethical approval was obtained from the Institutional Ethics Committee.

All available H&E-stained slides were retrieved and re-evaluated. Although the original cytological reports were issued prior to the 2023 Bethesda System, all cases were retrospectively reclassified according to this updated framework to ensure terminological consistency and facilitate comparability with current literature.

Perineural inflammation was defined as the presence of inflammatory cells circumferentially surrounding or infiltrating the perineurium of identifiable peripheral nerve twigs. Vasculitic-like changes were defined as dense perivascular lymphocytic inflammation involving small- to medium-sized intrathyroidal vessels, with focal mural inflammatory infiltration and endothelial swelling, in the absence of fibrinoid necrosis or thrombosis, thereby distinguishing this reactive pattern from a true necrotizing vasculitis. The fibrosis score was graded semi-quantitatively as follows: 0, no fibrosis; 1 (mild), focal thin fibrous septa involving less than 25% of the examined tissue; 2 (moderate), fibrous bands involving 25–50% of the tissue without distortion of the adjacent follicular architecture; and 3 (severe), broad, confluent fibrosis involving more than 50% of the tissue with architectural distortion. The SGT inflammation score reflected the extent of intraparenchymal granulomatous inflammation and was graded as Score 1 (mild), involving less than 30% of the parenchyma; Score 2 (moderate), involving 30–74%; and Score 3 (severe), involving 75% or more. Neutrophilic infiltration was recorded as present when neutrophils were identified within or at the periphery of granulomatous foci or forming microabscesses, a feature regarded as suggestive of early stage disease. Intragranulomatous necrosis was defined as central eosinophilic, karyorrhectic, or coagulative necrotic debris within granulomatous foci, distinguished from simple colloid degeneration by the presence of necrotic inflammatory cell debris and loss of the follicular epithelial outline. Oncocytic change was defined as the presence of follicular epithelial cells with abundant granular eosinophilic cytoplasm due to mitochondrial accumulation, involving non-granulomatous follicles adjacent to the inflammatory process.

Statistical analysis

Statistical analyses were performed using SPSS Statistics version 22.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics were presented as numbers and percentages for categorical variables and as mean, standard deviation, median, minimum, and maximum values for continuous variables. Comparisons between two independent groups were performed using Student's t-test when the assumption of normal distribution was met. Categorical variables were compared using the chi-square test or Fisher's exact test, as appropriate. A two-sided p-value of less than 0.05 was considered statistically significant.

RESULTS

Clinical and laboratory findings

The cohort consisted of 36 females (81.8%) and 8 males (18.2%), with a mean age of 48.1±8.9 years (range: 32–71). Among patients with available thyroid function data, 80% were euthyroid, 14.2% had subclinical hyperthyroidism, and 5.8% had overt hyperthyroidism.

FNAC was performed in 30 patients (68.2%). Cytological categories included nondiagnostic (6 cases, 20.0%), benign (9 cases, 30.0%), atypia of undetermined significance (3 cases, 10.0%), follicular neoplasm (5 cases, 16.7%), and suspicious for malignancy (7 cases, 23.3%). Only one case demonstrated cytological findings consistent with SGT. Surgical indications were multinodular goiter (28 cases, 63.6%), suspicious cytology (7 cases, 15.9%), follicular neoplasm (5 cases, 11.4%), coexisting parathyroid adenoma (3 cases, 6.8%), and refractory severe pain (1 case, 2.3%).

Macroscopic and histopathological findings

Bilateral pallor was observed in 14 cases (31.8%) and unilateral pallor in 11 cases (25.0%), whereas 19 cases (43.2%) showed no distinct macroscopic abnormality beyond nodular disease. The mean thyroid weight was 28.6±14.2 g (range: 6–76 g).

Microscopically, inflammation was classified as diffuse (17 cases, 38.6%), nodular (13 cases, 29.5%), or mixed (14 cases, 31.9%). SGT involvement was confined to parenchyma in 39 cases (88.6%), while 5 cases (11.4%) showed extension into areas of follicular nodular disease.

Granulomas were identified in all cases (100%), typically follicle-centered and coalescent. Multinucleated giant cells were present in all cases.

Neutrophilic infiltration, suggestive of early stage disease, was observed in 28 cases (63.6%).

Intragranulomatous necrosis was rare (2 cases, 4.5%) and limited to focal, submillimetric coagulative necrosis.

The extent of intraparenchymal inflammation was <30% in 17 cases (38.6%), 30–74% in 13 cases (29.5%), and ≥75% in 14 cases (31.9%). A semiquantitative scoring system was applied as follows: Score 1 (Mild, <30%), Score 2 (Moderate, 30–74%), and Score 3 (Severe, ≥75%), as defined for the purposes of this study.

Oncocytic changes were present in five cases (11.4%), frequently in association with chronic lymphocytic thyroiditis.

Perineural inflammatory infiltration was observed in 28 cases (63.6%), while vasculitic-like changes were present in 16 cases (36.4%) (Figure 1).

Fibrosis was a prominent feature, with severe fibrosis observed in 25 cases (56.8%), indicating advanced-stage disease.

Atrophic follicles within inflamed fibromyxoid stroma were identified in 28 cases (63.6%).

Inflammation in perithyroidal soft tissue was observed in 22 cases (50.0%) and edema in 20 cases (45.5%). These findings did not consistently overlap.

Follicular nodular disease (FND) was present in 39 cases (88.6%). Follicular adenoma was identified in six cases (13.6%) and chronic lymphocytic thyroiditis in seven cases (15.9%).

Concomitant malignancy was identified in 13 cases (29.5%), including classic subtype subcentimeter papillary thyroid carcinoma (7 cases, 15.9%), minimally invasive encapsulated follicular variant papillary carcinoma (3 cases, 6.8%), tall cell subtype subcentimeter papillary thyroid carcinoma (2 cases, 4.5%), and infiltrative follicular subtype papillary carcinoma (1 case, 2.3%) (Figure 2).

Comparison of clinicopathological parameters between cases with and without malignancy revealed no statistically significant differences in sex, age, thyroid weight, fibrosis score, SGT inflammation score, neutrophilic infiltration, perineural inflammation, intragranulomatous necrosis, lymph node status, chronic lymphocytic thyroiditis, oncocytic change, or vasculitic-like changes (all p>0.05; Table 1).

DISCUSSION

SGT is traditionally regarded as a self-limited inflammatory disorder with well-defined clinicopathological features[1]. However, the findings of the present study suggest that SGT may have broader diagnostic and pathophysiological implications than previously appreciated. By systematically evaluating 44 thyroidectomy specimens, we identified three key aspects that merit particular attention: (i) the potential role of neural inflammation in pain pathogenesis, (ii) the unexpectedly high rate of coexisting thyroid carcinoma, and (iii) significant limitations of preoperative cytological diagnosis.

The most striking finding was the high frequency of peripheral nerve inflammation, observed in 63.6% of cases. Pain is the most prominent clinical feature of subacute granulomatous thyroiditis. The pain occurring in the thyroid gland is generally attributed to stretching of the thyroid capsule due to inflammation and enlargement of the gland[2-4]. However, this explanation does not fully account for cases with severe

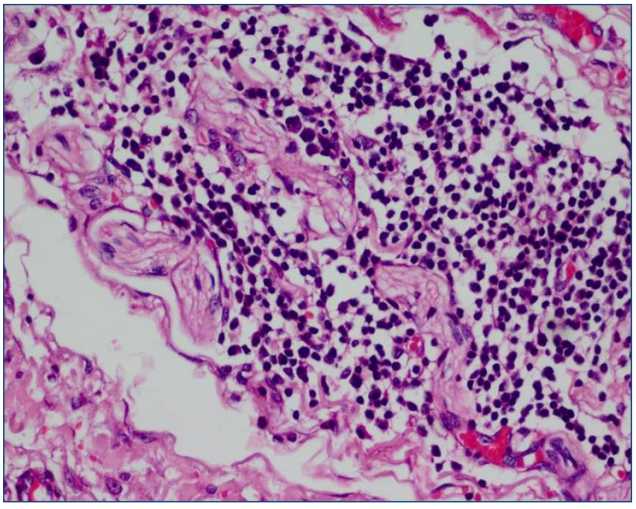


Figure 1. Lymphoplasmacytic inflammatory cell infiltration within and around peripheral nerves (hematoxylin & eosin, ×400).

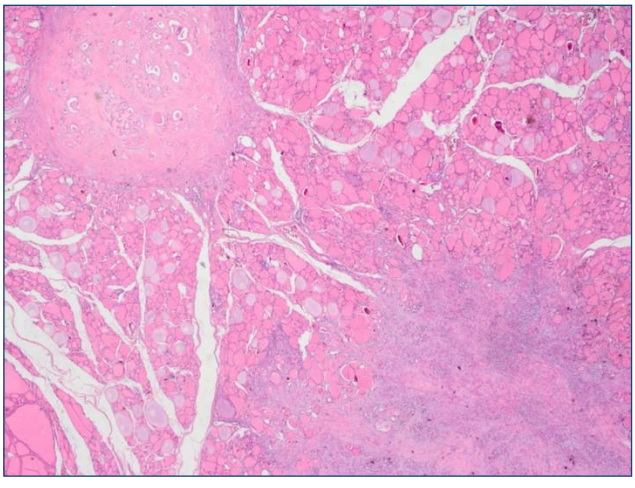


Figure 2. Upper left: subcentimeter classic subtype of papillary thyroid carcinoma. Lower right: focus of subacute granulomatous thyroiditis (hematoxylin & eosin, ×20).

Table 1. Comparison of clinicopathological features according to malignancy status.

Variable	Malignancy present (n=13)	No malignancy (n=31)	p-value
Sex			
Female	11 (84.6%)	25 (80.6%)	1.000*
Male	2 (15.4%)	6 (19.4%)	
Age (years)	48.5±5.7	48.0±10.0	0.813**
Thyroid weight (g)	29.1±13.2	28.3±15.0	0.878**
Fibrosis score			
0	1 (7.7%)	3 (9.7%)	0.775*
1	3 (23.1%)	4 (12.9%)	
2	3 (23.1%)	5 (16.1%)	
3	6 (46.2%)	19 (61.3%)	
SGT inflammation score†			
Score 1 (Mild, <30%)	6 (46.2%)	11 (35.5%)	0.713*
Score 2 (Moderate, 30–74%)	4 (30.8%)	9 (29.0%)	
Score 3 (Severe, ≥75%)	3 (23.1%)	11 (35.5%)	
Neutrophilic infiltration: present	8 (61.5%)	20 (64.5%)	1.000*
Perineural inflammation: present	8 (61.5%)	20 (64.5%)	1.000*
Intragranulomatous necrosis: present	1 (7.7%)	1 (3.2%)	0.508*
Incidental lymph node: present	7 (53.8%)	10 (32.3%)	0.309*
Chronic lymphocytic thyroiditis: present	4 (30.8%)	3 (9.7%)	0.170*
Oncocytic change: present	3 (23.1%)	2 (6.5%)	0.144*
Vasculitic-like changes: present	5 (38.5%)	11 (35.5%)	1.000*

*Fisher's exact test; **Student's t-test. †SGT inflammation score: Score 1: <30%; Score 2: 30–74%; Score 3: ≥75% parenchymal involvement. Continuous variables: mean±SD; categorical variables: n (%). SGT: subacute granulomatous thyroiditis.

pain in the absence of marked enlargement[2]. Our findings provide histopathological support for an alternative or complementary mechanism: direct neural involvement by inflammatory infiltrates. Perineural inflammation has been implicated in pain generation in various inflammatory and neoplastic conditions[5,6], yet its role in thyroid pathology remains underexplored. Neural involvement has rarely been addressed in previous histopathological studies of SGT, making this observation a potentially underrecognized feature of the disease and warranting further investigation through clinicopathological correlation studies. However, we were unable to perform preoperative pain assessments due to the time elapsed since the surgeries and our inability to reach a significant portion of the patients. Therefore, it was not possible for us to establish a correlation between pain and the severity of perineural inflammation.

Another notable finding is the relatively high prevalence of coexisting thyroid malignancies (29.5%), predominantly subcentimeter papillary thyroid carcinomas. At first glance, this rate appears to exceed those reported in the literature and could raise the question of a potential pathogenetic link between SGT and thyroid carcinoma[7]. However, several lines of evidence argue against a causal relationship. First, the majority of detected malignancies were subcentimeter papillary thyroid carcinomas, many of which were incidental findings. Second, no histopathological transition or spatial association suggesting progression from inflammatory lesions to carcinoma was observed. Third, well-established risk modifiers such as multinodular goiter and chronic lymphocytic thyroiditis were frequently present in the background thyroid tissue. Taken together, we interpret this finding as most likely reflecting incidental detection related to extensive histopathological sampling inherent to total thyroidectomy specimens, rather than a true pathogenetic link. Notably, the observed malignancy rate is comparable to the incidental carcinoma rates routinely encountered in our institution[8,9]. In this regard, SGT may paradoxically increase the detection of incidental carcinomas by prompting more meticulous sampling due to its irregular and suspicious gross appearance.

From a surgical pathology perspective, these findings also have important practical implications. Because SGT is rarely encountered in thyroidectomy specimens, its histopathological spectrum may be unfamiliar to pathologists who infrequently evaluate such cases. Moreover, the irregular gross appearance and extensive inflammatory changes may mimic thyroid malignancy or obscure coexisting neoplastic lesions. Therefore, careful gross examination, adequate histopathological sampling, and thorough microscopic evaluation are essential to establish the correct diagnosis and to avoid overlooking incidental thyroid lesions. The relatively high frequency of incidental carcinoma observed in our series is most likely attributable to extensive histopathological sampling rather than a true biological association. Accordingly, these findings should not be interpreted as evidence that SGT represents a preneoplastic condition.

The limited diagnostic utility of FNAC is a further important finding. Despite being performed in 68.2% of cases, FNAC correctly suggested SGT in only one instance. This low diagnostic yield is likely because FNAC is typically directed toward dominant or suspicious nodules rather than the inflamed parenchyma, thereby missing characteristic granulomatous lesions. Among patients undergoing FNAC, seven were reported as suspicious for malignancy (Bethesda V) and five as follicular neoplasm (Bethesda IV), directly precipitating surgery. These data highlight a critical limitation of FNA in this setting and suggest that SGT may be substantially underrecognized preoperatively, particularly in the presence of coexisting nodular disease[10]. Re-evaluation of cytological specimens or repeat FNAC may help avoid unnecessary surgical intervention[11,12].

Thyroidectomy is rarely indicated for SGT; NSAIDs are first-line therapy, with corticosteroids reserved for severe or refractory symptoms[13]. Nevertheless, a small number of cases requiring surgery because of severe pain attributable to SGT have been reported in the literature. These reports suggest that thyroidectomy may be considered in selected patients with medically refractory disease and persistent symptoms despite appropriate medical treatment[14,15]. Interestingly, one patient in our cohort underwent surgery with a preoperative diagnosis of SGT after experiencing persistent symptoms despite intermittent corticosteroid therapy for approximately two years. This case further supports the notion that surgical intervention may occasionally have a role in the management of treatment-resistant SGT.

Vasculitic-like changes were identified in 36.4% of cases. Previous reports have suggested that vasculitic processes may play a role in the pathogenesis of SGT, particularly in association with systemic vasculitic disorders such as granulomatosis with polyangiitis, Takayasu arteritis, and Henoch-Schönlein purpura[16,17]. None of our patients had documented systemic vasculitis; nonetheless, these findings suggest that vascular immune-mediated mechanisms may contribute to the inflammatory process in a subset of patients.

A major strength of the present study is the evaluation of a relatively large series of surgically treated SGT cases, a condition that is only rarely managed by thyroidectomy. Comprehensive histopathological re-review of all available specimens enabled systematic assessment of both classical and less commonly described features, including perineural inflammation and vasculitic-like changes, as well as coexisting benign and malignant thyroid lesions.

Nevertheless, several limitations should be acknowledged. The retrospective design and restriction to surgically treated cases introducing inherent selection bias—the observed frequencies, particularly regarding coexisting malignancy, may not reflect true prevalence in the general SGT population. The absence of a control group limited assessment of the specificity of neural and vasculitic findings. Pain severity could not be retrospectively quantified, precluding direct correlation with perineural inflammation. The relatively limited sample size may have reduced statistical power for subgroup analyses.

In conclusion, the frequent presence of perineural inflammation represents a previously underrecognized histopathological finding that may contribute to the pathogenesis of thyroid pain in SGT. Our findings further suggest that the apparent association with thyroid carcinoma is most likely attributable to incidental detection resulting from extensive histopathological sampling rather than a true biological relationship. From a diagnostic perspective, SGT may present significant challenges because its gross appearance can mimic malignancy and obscure coexisting lesions. Therefore, careful gross examination, adequate histopathological sampling, and thorough microscopic evaluation of thyroidectomy specimens are essential to ensure accurate diagnosis and to identify incidental coexisting thyroid neoplasms.

ETHICAL APPROVAL

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki and applicable data privacy legislation. Ethical approval was obtained from the Non-Interventional Clinical Research Ethics Committee of Recep Tayyip Erdoğan University Faculty of Medicine (approval number: 2021/188, date: 11/11/2021).

AUTHORS' CONTRIBUTIONS

MYA: Data curation, Writing – original draft. **HG:** Conceptualization, 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. Engkakul P, Mahachoklertwattana P, Poomthavorn P. Eponym: de Quervain thyroiditis. Eur J Pediatr. 2011;170(4):427-31. https://doi.org/10.1007/s00431-010-1306-4
2. Rotondi M, Capelli V, Locantore P, Pontecorvi A, Chiovato L. Painful Hashimoto's thyroiditis: myth or reality? J Endocrinol Invest. 2017;40(8):815-8. https://doi.org/10.1007/s40618-017-0655-5
3. Furszyfer J, Kurland LT, Woolner LB, Elveback LR, McConahey WM. Hashimoto's thyroiditis in Olmsted County, Minnesota, 1935 through 1967. Mayo Clin Proc. 1970;45(8):586-96.
4. Li JH, Daniels GH, Barbesino G. Painful subacute thyroiditis is commonly misdiagnosed as suspicious thyroid nodular disease. Mayo Clin Proc Innov Qual Outcomes. 2021;5(2):330-7. https://doi.org/10.1016/j.mayocpiqo.2020.12.007
5. Papa J, Pucchio AC, Schneider M, Wang A. Perineural inflammation as a novel feature in lichen sclerosus: a case series of histologic and clinical features. Am J Dermatopathol. 2024;46(5):287-91. https://doi.org/10.1097/dad.0000000000002640
6. Mousa SA, Shaqura M, Mohamed D, Hong X, Krug SM, Metwally EY, et al. Tumor-Macrophage-Nerve interactions drive neuroinflammation and neuropathic pain in prostate cancer perineural invasion. Brain Behav Immun. 2026;133:106280. https://doi.org/10.1016/j.bbi.2026.106280
7. Gül N, Üzüm AK, Selçukbiricik ÖS, Yegen G, Tanakol R, Aral F. Prevalence of papillary thyroid cancer in subacute thyroiditis patients may be higher than it is presumed: retrospective analysis of 137 patients. Radiol Oncol. 2018;52(3):257-62. https://doi.org/10.2478/raon-2018-0027
8. Kefeli M, Gucer H, Durmuş ET, Atmaca A, Colak R, Mete O. Prevalence of papillary thyroid carcinoma is significantly higher in graves disease with synchronous thyroid nodules. Turk Patoloji Derg. 2024;40(3):190-5. https://doi.org/10.5146/tjpath.2024.13650
9. Semerci O, Gucer H. The significance of unsampled microscopic thyroid carcinomas in multinodular goiter. Endocr Pathol. 2023;34(1):119-28. https://doi.org/10.1007/s12022-022-09743-z
10. Lamichaney R, Sherpa M, Das D, Bhutia CT, Laishram S. Fine-needle aspiration of de quervain's thyroiditis (subacute granulomatous thyroiditis): a cytological review of 20 cases. J Clin Diagn Res. 2017;11(8):Ec09-11. https://doi.org/10.7860/jcdr/2017/26054.10355
11. Duininck TM, Heerden JA, Fatourechi V, Curlee KJ, Farley DR, Thompson GB, et al. de Quervain's thyroiditis: surgical experience. Endocr Pract. 2002;8(4):255-8. https://doi.org/10.4158/ep.8.4.255
12. Ranganath R, Shaha MA, Xu B, Migliacci J, Ghossein R, Shaha AR. de Quervain's thyroiditis: a review of experience with surgery. Am J Otolaryngol. 2016;37(6):534-7. https://doi.org/10.1016/j.amjoto.2016.08.006
13. Tabassom A, Chippa V, Edens MA. De quervain thyroiditis. StatPearls. StatPearls Publishing Copyright © 2026, StatPearls Publishing LLC; 2026.
14. Mazza E, Quaglino F, Suriani A, Palestini N, Gottero C, Leli R, et al. Thyroidectomy for painful thyroiditis resistant to steroid treatment: three new cases with review of the literature. Case Rep Endocrinol. 2015;2015:138327. https://doi.org/10.1155/2015/138327
15. Szczepanek-Parulska E, Zybek A, Biczysko M, Majewski P, Ruchała M. What might cause pain in the thyroid gland? Report of a patient with subacute thyroiditis of atypical presentation. Endokrynol Pol. 2012;63(2):138-42.
16. Mukae H, Furusyo N, Murata M, Ogawa E, Kainuma M, Shimizu M, et al. A case of granulomatosis with polyangiitis preceded by subacute thyroiditis. Clin Case Rep. 2015;3(3):139-44. https://doi.org/10.1002/ccr3.181
17. Ohta Y, Ohya Y, Fujii K, Tsuchihashi T, Sato K, Abe I, et al. Inflammatory diseases associated with Takayasu's arteritis. Angiology. 2003;54(3):339-44. https://doi.org/10.1177/000331970305400310

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260877

Potentially avoidable computed tomography after positive ultrasonography in appendicitis

Serdar Derya[1]*, İpek Ergün[2], Ahmet Arda Altuner[2], Ahmet Ege Büyükfirat[2], Ahsen Öztürk[2], Ali Efe Aktürk[2], Aram Miran Yaruk[2], Ahmet Deniz Topal[2], Yağmur Irmak[2], Zehra Keser[2]

SUMMARY

OBJECTIVE: The aim of this study was to determine the frequency of computed tomography performed despite a positive ultrasonography in histopathologically confirmed appendicitis (potentially avoidable computed tomography), and to assess its effects on radiation exposure, cost, and emergency department length of stay.

METHODS: This single-center retrospective cohort study included 887 consecutive patients who underwent appendectomy for suspected appendicitis (January 2023–December 2025). Potentially avoidable computed tomography was defined as computed tomography following a positive ultrasonography result. As length of stay was non-normal, imaging strategies were compared using the Kruskal-Wallis test, and length of stay predictors were modeled using median (quantile) regression, with a log-link gamma generalized linear model as a sensitivity analysis. Imaging cost used national tariffs; radiation was estimated as ~8 mSv per abdominal computed tomography with age-specific BEIR VII multipliers.

RESULTS: Computed tomography was used in 87.6% (777/887) of cases. Of the 521 patients who underwent ultrasonography, the results were positive in 35.5%, inconclusive in 30.7%, and negative in 33.8% of cases. Potentially avoidable computed tomography was performed in 81/777 computed tomography examinations (10.4%). Notably, 43.8% (81/185) of patients with positive ultrasonography results subsequently underwent computed tomography scan, corresponding to ~648 mSv of avoidable radiation exposure and an associated cost of 26,667 TL (~US$811), with 49.4% of the affected patients being children. Length of stay did not differ across strategies (p=0.748), and computed tomography was not an independent length of stay predictor (+7.2 min; p=0.23), whereas business-hour arrival was (+15.3 min; p<0.001).

CONCLUSION: Computed tomography remained frequent even after ultrasonography established the diagnosis, with no effect on emergency department length of stay, generating avoidable radiation and costs concentrated in children. Reducing computed tomography after a positive ultrasonography is a feasible opportunity to improve resource utilization while preserving patient safety.

KEYWORDS: Appendicitis. Emergency service. Radiation dosage. Tomography. Ultrasonography.

INTRODUCTION

Acute appendicitis is one of the most common surgical emergencies worldwide; however, its diagnosis is challenging. Because clinical findings and laboratory tests alone are insufficient, scoring systems combining clinical and imaging data are widely used to support diagnosis and to distinguish complicated from uncomplicated disease[1]. Misdiagnosis and unnecessary appendectomy remain notable, particularly in atypical presentations, and diagnostic delay is associated with perforation and intra-abdominal abscess[2]. Imaging that provides little or no patient benefit lowers care quality and burdens the health system[3]. Reducing low-value imaging is therefore a principal quality-improvement goal in emergency medicine and a core theme of the Choosing Wisely initiative[4,5].

Ultrasonography (US), being radiation-free and widely available, is the first-line modality, especially in children and non-obese adults[6,7]. However, many US examinations are inconclusive, prompting further imaging or observation[8]. Computed tomography (CT), owing to its high diagnostic accuracy, is an important complementary modality when clinical evaluation is uncertain[9,10]. In some cases, CT is still performed even after US has clearly established appendicitis; we defined such examinations as potentially avoidable CT (PA-CT).

Imaging overuse is a particular focus of Choosing Wisely, which recommends limiting unnecessary or repeated examinations[5]. Abdominal CT delivers about 8 mSv per examination, important in children for lifetime cancer risk; both the ICRP 103 recommendations and the EuroSafe Imaging

[1]Inonu University, Faculty of Medicine, Department of Emergency Medicine – Malatya, Türkiye.
[2]Inonu University, Faculty of Medicine – Malatya, Türkiye.
*Corresponding author: serdar.derya@inonu.edu.tr

Conflicts of interest: the authors declare there is no conflicts of interest. Funding: this article was supported by the Inonu University Scientific Research Projects Unit (TSA-2026-4834).
Received on June 21, 2026. Accepted on July 25, 2026.
Scientific Editor: José Maria Soares Júnior

initiative emphasize keeping dose as low as reasonably achievable (ALARA)[11-13].

Although the diagnostic performance of US and CT is well-defined, studies evaluating how imaging strategies affect resource use and ED operations in histopathologically confirmed cohorts are limited. This is consistent with Chidiac et al., who emphasized reducing radiation exposure in children with suspected appendicitis[14], and with the systematic review and meta-analysis by Castro-Luna et al. on US, CT, and MRI for acute appendicitis in children and adolescents[15].

We aimed to determine the frequency of PA-CT in a three-year histopathologically confirmed cohort, its associated radiation burden and cost, and its relationship with ED LOS. Cross-sectional imaging may be clinically necessary to exclude complications or delineate atypical anatomy; therefore, the PA-CT definition reflects potential optimization rather than a direct indicator of care.

METHODS

Study design and ethics

This single-center retrospective cohort study was conducted at a tertiary emergency service (January 2023–December 2025), approved by the institutional research ethics committee (no. 2026/9399), and conducted in accordance with the Declaration of Helsinki. The requirement for individual informed consent was waived owing to the retrospective nature of the study. Reporting followed the STROBE statement[16].

Participants

All patients presenting with suspected acute appendicitis who underwent emergency appendectomy with complete ED records were included (n=887) and classified by imaging strategy: US only (n=110), US followed by CT (n=411), and direct CT without US (n=366).

Definitions and variables

A positive US result was defined as a non-compressible, blind-ending tubular structure in the right lower quadrant with an outer diameter >6 mm (with or without wall thickening, periappendiceal fat changes, or an appendicolith). Examinations that did not meet these criteria were classified as negative or inconclusive when the appendix could not be visualized on the images. PA-CT was performed after a positive US result (examinations as potentially avoidable CT). This procedural definition does not include the CT scan requested to rule out appendicitis complications such as suspected perforation, abscess, or retrocecal anatomy. LOS was defined as the interval from triage to the operating room transfer. Imaging costs used national tariffs (CT 329.22 TL; US 78.33 TL), also expressed in US$ (32.9 TL/US$). Radiation was estimated as 8 mSv per abdominal CT[17] and age-specific BEIR VII risk multipliers[12,13].

Statistical analysis

Continuous variables are presented as median (interquartile range), and categorical variables are presented as n (%). As LOS was non-normal (Shapiro-Wilk W=0.82, p<0.001), groups were compared using Kruskal-Wallis and Mann-Whitney tests, and LOS predictors were modeled with median (quantile) regression, with a log-link gamma generalized linear model (GLM) as a sensitivity analysis. Consultation duration was excluded as a within-window subcomponent of LOS. The model included 887 patients after listwise deletion of missing values (mainly leukocyte count, n=56). Categorical comparisons used the χ^2 or Fisher exact test; baseline comparisons were descriptive. Analyses were performed using Python 3.12 (SciPy, Statsmodels)[18]; p<0.05 was considered significant.

RESULTS

Patient characteristics

Of the 887 patients, 57.2% were male and 38.8% were children (<18 years); the median age was 21 years (IQR, 13–41). The negative appendectomy rate was 0.23% (2/887). The baseline characteristics of the patients according to the imaging strategy are presented in Table 1. After Bonferroni correction (p<0.004), differences in age, proportion of pediatric patients, sex, symptom duration, right lower quadrant tenderness, and fever persisted.

Imaging strategies and potentially avoidable computed tomography

CT was used in 87.6% of cases (n=777; Figure 1). Of the 521 patients who underwent US, the results were positive in 185 (35.5%), inconclusive in 160 (30.7%), and negative in 176 (33.8%); complementary CT after inconclusive and negative US results was common. PA-CT was performed in 81/777 CT examinations (10.4%); 43.8% of patients nonetheless underwent CT. Compared with other CT patients, PA-CT patients had similar C-reactive protein (1.7 vs. 2.0 mg/L, p=0.829), leukocytes (13.4 vs. 13.0×10^3/µL, p=0.597), and neutrophils (80.0 vs. 79.2%, p=0.285), and lower fever frequency (9.9 vs. 15.8%, p=0.001). In 772/777 CT examinations, the findings were consistent with the histopathological results. As all patients

Table 1. Baseline characteristics by imaging strategy (n=887).

Variable	Overall (n=887)	US only (n=110)	US+CT (n=411)	CT only (n=366)	PA-CT (n=81)	Other CT (n=696)	p-value
Age, years, median (IQR)	21 (13–41)	13 (10–18)	16 (11–31)	33 (22–52)	18 (12–43)	24 (15–44)	<0.001
Pediatric (<18 y), n (%)	344 (38.8)	78 (70.9)	239 (58.2)	27 (7.4)	40 (49.4)	226 (32.5)	<0.001
Male sex, n (%)	507 (57.2)	81 (73.6)	223 (54.3)	203 (55.5)	42 (51.9)	384 (55.2)	<0.001
Symptom duration, h	24 (24–24)	24 (24–25)	24 (15–24)	24 (24–24)	23 (2–24)	24 (24–24)	0.002
RLQ tenderness, n (%)	527 (59.4)	51 (46.4)	219 (53.3)	257 (70.2)	58 (71.6)	418 (60.1)	<0.001
Rebound tenderness, n (%)	188 (21.2)	14 (12.7)	80 (19.5)	94 (25.7)	20 (24.7)	154 (22.1)	0.007
Fever ≥38°C, n (%)	150 (16.9)	32 (29.1)	58 (14.1)	60 (16.4)	8 (9.9)	110 (15.8)	<0.001
WBC, $\times 10^3/\mu L$	13.0 (10.5–15.9)	14.0 (11.1–17.9)	13.3 (10.3–16.7)	12.7 (10.5–15.0)	13.4 (10.6–16.6)	13.0 (10.4–15.6)	0.009
Neutrophil%	79.2 (69.7–85.0)	78.0 (70.5–85.1)	80.4 (69.0–86.0)	77.0 (70.0–83.7)	80.0 (72.8–86.0)	79.2 (69.6–84.8)	0.060
CRP, mg/L	2.0 (0.4–6.2)	1.4 (0.1–5.4)	1.6 (0.1–6.8)	2.7 (0.6–6.0)	1.7 (0.5–6.7)	2.0 (0.5–6.2)	0.034
Business-hours arrival, n (%)	439 (49.5)	57 (51.8)	200 (48.7)	182 (49.7)	40 (49.4)	342 (49.1)	0.835
ED LOS, h	5.77 (5.25–6.18)	5.70 (5.30–6.11)	5.75 (5.22–6.17)	5.77 (5.27–6.21)	5.9 (5.4–6.2)	5.8 (5.2–6.2)	0.748
Negative appendectomy, n (%)	2 (0.23)	0 (0.0)	0 (0.0)	2 (0.55)	0 (0.0)	2 (0.3)	0.533

Values are presented as median (IQR) or number (%). p-values from the Kruskal-Wallis test (continuous) and χ^2 test (categorical), with Fisher's exact test for expected counts <5. Comparisons were descriptive; under a Bonferroni threshold (p<0.004), age, proportion of pediatric patients, sex, symptom duration, RLQ tenderness, and fever were significant. RLQ: right lower quadrant; LOS: length of stay; WBC: white blood cell; CRP: C-reactive protein; US: ultrasonography; CT: computed tomography; PA-CT: potentially avoidable computed tomography; IQR: interquartile range; ED: emergency department. PA-CT vs. other CT comparisons were post-hoc; under Bonferroni correction (4 tests, threshold p<0.0125), only fever frequency remained significant.

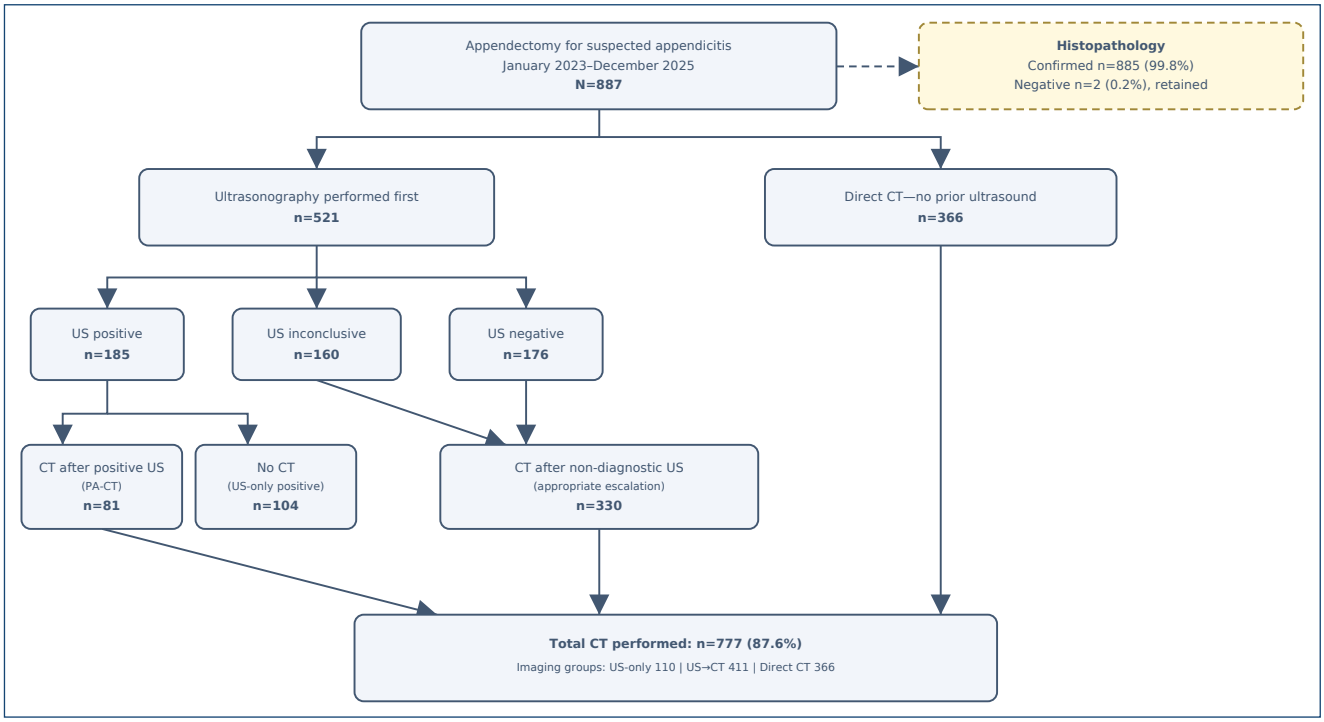


Figure 1. Patient flow and imaging pathway.

underwent surgery, the specificity and negative predictive values could not be calculated.

Emergency department length of stay

The median LOS was 5.77 h (IQR 5.25–6.18) and did not differ across strategies (5.70, 5.75, 5.77 h; p=0.748); CT-versus-no-CT (p=0.627) and age group (p=0.444) comparisons were also non-significant. In median regression (n=828), CT was not independently associated with LOS (+7.2 min; 95%CI -4.5 to 18.8; p=0.23), whereas business-hours arrival was (+15.3 min; 95%CI 8.2–22.3; p<0.001); the gamma GLM concurred [CT exp(β)=1.02, p=0.34; business hours exp(β)=1.13, p<0.001]. No clinical or laboratory variable was independently associated with LOS (all p>0.10), and the model's explanatory power was limited (gamma pseudo-R^2≈0.09). Among patients with positive US results, the median LOS was 5.88 h with PA-CT versus 5.67 h without CT (p=0.058).

Cost and radiation

The total imaging cost was 296,614 TL (56.5% from US-guided CT, 40.6% from direct CT). PA-CT alone corresponded to 26,667 TL (~US$811) and 648 mSv of potentially avoidable radiation exposure, respectively. The PA-CT group had a median age of 18 years, and 49.4% of the patients were children (Table 2). The total CT radiation was ~6,216 mSv, of which 2,128 mSv (34.2%) was attributable to the children.

DISCUSSION

In 887 patients with histopathologically confirmed appendicitis, CT was performed despite a positive US in 10.4% of cases, with 648 mSv of potentially avoidable radiation and ~US$811, with approximately half of the affected patients being children. These post-US CT examinations generated a considerable radiation and cost burden but had no significant effect on ED LOS; the rationale for reducing them therefore relates to preventing unnecessary radiation and using health resources more efficiently rather than to patient flow[19,20].

The fact that 43.8% of positive US patients still underwent CT may reflect the exclusion of complications, anatomical evaluation for surgical planning, established practice, and medicolegal or defensive medicine considerations, together with limited confidence in operator-dependent US. In multivariable analyses, LOS was independent of the imaging strategy and was driven more by operational factors; business-hour arrival was associated with LOS, whereas CT was not. A longer LOS during business hours likely reflects a higher daytime ED census and competition for imaging and operating room capacity rather than imaging choice and is hypothesis-generating. Additional CT after a positive US provides no measurable advantage to patient flow.

The 35.5% positive US rate may reflect the operator-dependent nature of US and the classification of non-visualized appendices as inconclusive findings. In the positive US subgroup, the trend toward longer LOS in PA-CT patients did not reach significance (p=0.058), consistent with the absence of a clear operational benefit.

Radiation findings are particularly important in children, in whom longer life expectancy and higher radiosensitivity make long-term risks more pronounced than in adults[21,22]. Avoiding unnecessary CT when US adequately supports the diagnosis is therefore feasible for radiation safety, and low-dose CT protocols offer additional risk reduction when CT is needed[9]. A pediatric appendicitis algorithm with a US-first approach and an option for next-day surgery clinic follow-up in uncertain cases reduced CT use from 13.8 to 6%, without increasing US use, hospital admission, surgical consultation, ED return within 7 days, or LOS[23].

Table 2. Imaging cost and radiation by age group.

Age group	CT, n (%)	Total mSv	BEIR VII ×	PA-CT, n	Avoidable mSv	Avoidable cost, TL
0–5 year	27 (3.5)	216	3.0×	4	32	1,317
6–12 year	129 (16.6)	1,032	2.5×	17	136	5,597
13–17 year	110 (14.2)	880	2.0×	19	152	6,255
18–40 year	291 (37.5)	2,328	1.0× (ref)	16	128	5,268
41–60 year	138 (17.8)	1,104	0.6×	18	144	5,926
>60 year	82 (10.6)	656	0.3×	7	56	2,305
Total	777 (100)	6,216	–	81	648	26,667

Avoidable cost=PA-CT count×CT tariff (329.22 TL); total 26,667 TL (≈US$811). CT: computed tomography; PA-CT: potentially avoidable computed tomography; TL: Turkish Lira.

The very low negative appendectomy rate (0.23%) may reflect high imaging use and a selective surgical approach, although histopathological criteria and local practice may also contribute; this should not be generalized to other centers. Importantly, because PA-CT denotes CT performed after an already positive US, it identifies the subset in which CT is least likely to alter the operative decision; therefore, reducing CT in this group is unlikely to compromise the low negative appendectomy rate, which reflects the overall imaging-and-selective-surgery strategy rather than the incremental post-US scan alone.

LIMITATIONS

PA-CT is a procedural definition. Complicated disease findings (perforation, abscess, plastron) were not systematically recorded; therefore, post-US CT may have been justified in some instances. CT was ordered by both emergency physicians and surgeons; however, the ordering clinician was not recorded in every case, limiting the evaluation of the decision-making process. The retrospective, single-center design and inclusion of only operated patients limit the generalizability of our findings to some extent. Radiation was estimated from reference doses rather than patient-based dosimetry, and the cost analysis covered only direct imaging costs, not indirect costs (equipment depreciation, personnel, and ED crowding). Finally, the models explained only a limited portion of the variation in LOS.

CONCLUSION

Despite a US-based diagnosis of appendicitis, CT remains frequently used, with no significant effect on ED LOS, and is associated with potentially avoidable radiation and costs concentrated in children. Reconsidering the use of CT after a positive US offers a feasible opportunity to optimize imaging practices. Prospective multicenter studies that capture complicated disease features and clinicians making imaging decisions are needed to confirm these observations and clarify the clinical rationale for post-US CT.

ETHICS APPROVAL

Institutional research ethics committee No. 2026/9399.

AUTHORS' CONTRIBUTIONS

SD: Conceptualization, Formal Analysis, Investigation, Methodology, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. İE: Data curation, Investigation, Resources, Validation, Writing – original draft. **AAA:** Data curation, Investigation, Resources, Validation, Writing – original draft. **AEB:** Data curation, Investigation, Resources, Validation, Writing – original draft. **AÖ:** Data curation, Investigation, Resources, Validation, Writing – original draft. **AEA:** Data curation, Investigation, Resources, Validation, Writing – original draft. **AMY:** Data curation, Investigation, Resources, Validation, Writing – original draft. **ADT:** Data curation, Investigation, Resources, Validation, Writing – original draft. **YI:** Data curation, Investigation, Resources, Validation, Writing – original draft. **ZK:** Data curation, Investigation, Resources, Validation, 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. Bom WJ, Scheijmans JCG, Salminen P, Boermeester MA. Diagnosis of uncomplicated and complicated appendicitis in adults. Scand J Surg. 2021;110:170-9.
2. Graff L, Russell J, Seashore J, Tate J, Elwell A, Prete M, et al. False-negative and false-positive errors in abdominal pain evaluation: failure to diagnose acute appendicitis and unnecessary surgery. Acad Emerg Med. 2000;7:1244-55.
3. Kjelle E, Brandsæter IØ, Andersen ER, Hofmann BM. Cost of low-value imaging worldwide: a systematic review. Appl Health Econ Health Policy. 2024;22:485-501.
4. Venkatesh AK, Schuur JD. A "top five" list for emergency medicine: a policy and research agenda for stewardship to improve the value of emergency care. Am J Emerg Med. 2013;31:1520-4.
5. Rao VM, Levin DC. The overuse of diagnostic imaging and the Choosing Wisely initiative. Ann Intern Med. 2012;157:574-6.
6. Debnath J, George RA, Ravikumar R. Imaging in acute appendicitis: what, when, and why? Med J Armed Forces India. 2017;73:74-9.
7. Roberts K, Moore H, Raju M, Gent R, Piotto L, Taranath A, et al. Diagnostic ultrasound for acute appendicitis: the gold standard. J Pediatr Surg. 2024;59:235-9.
8. Lehmann B, Koeferli U, Sauter TC, Exadaktylos A, Hautz WE. Diagnostic accuracy of a pragmatic, ultrasound-based approach to adult patients with suspected acute appendicitis in the ED. Emerg Med J. 2022;39:931-6.
9. Asim SJ, Masood Z, Soomro E, Qureshi AS. Diagnostic accuracy of low dose CT-scan abdomen in patients with clinical features of acute appendicitis. Pak J Med Sci. 2024;40:2052-5.

10. Schildberg CW, Reissig K, Hunger R, Paasch C, Stillger R, Mantke R. Diagnostic, therapy and complications in acute appendicitis of 19,749 cases based on routine data: a retrospective multicenter observational study. J Clin Med. 2022;11:4495.
11. Damilakis J, Granata C, Kotter E, Brkljacic B. Balancing patients' benefits and risks in computed tomography: a European Society of Radiology 'EuroSafe Imaging' viewpoint. Eur Radiol. 2025;36(2):1363-5. https://doi.org/10.1007/s00330-025-11911-8
12. Valentin J. The 2007 recommendations of the International Commission on Radiological Protection. ICRP Publication 103. Ann ICRP. 2007;37:1-332.
13. National Research Council. Health risks from exposure to low levels of ionizing radiation: BEIR VII Phase 2. Washington: National Academies Press; 2006.
14. Chidiac C, Issa O, Garcia AV, Rhee DS, Slidell MB. Failure to significantly reduce radiation exposure in children with suspected appendicitis in the United States. J Pediatr Surg. 2024;59:161701.
15. Castro-Luna DI, Porras-Hernandez JD, Flores-Garcia JA, Dies-Suarez P, Servin-Martinez MF, Pierdant-Perez M. Contemporary ultrasound, computed tomography, or magnetic resonance imaging for acute appendicitis diagnosis in children and adolescents: systematic review and meta-analysis. Pediatr Radiol. 2025;55:1448-64.
16. Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453-7. https://doi.org/10.1016/S0140-6736(07)61602-X
17. Khajmi H, Nouh FA, Tounsi A, Touti R. Evaluation of X-ray doses in pediatrics multislice computed tomography. Radioprotection. 2023;58(4):303-8. https://doi.org/10.1051/radiopro/2023029
18. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. Proc 9th Python Sci Conf. 2010:92-6.
19. Wang RC, Fahimi J, Dillon DG, Shyy W, Mongan JT, McCulloch CE, et al. Effect of an ultrasound-first clinical decision tool in emergency department patients with suspected nephrolithiasis: a randomized trial. Am J Emerg Med. 2022;60:164-70.
20. Ramarajan N, Krishnamoorthi R, Barth R, Ghanouni P, Mueller C, Dannenburg B, et al. An interdisciplinary initiative to reduce radiation exposure: evaluation of appendicitis in a pediatric emergency department with clinical assessment supported by a staged ultrasound and computed tomography pathway. Acad Emerg Med. 2009;16:1258-65.
21. Miglioretti DL, Johnson E, Williams A, Greenlee RT, Weinmann S, Solberg LI, et al. The use of computed tomography in pediatrics and the associated radiation exposure and estimated cancer risk. JAMA Pediatr. 2013;167:700-7.
22. Brenner D, Elliston C, Hall E, Berdon W. Estimated risks of radiation-induced fatal cancer from pediatric CT. AJR Am J Roentgenol. 2001;176:289-96.
23. Ryan S, Dudley NC, Schunk JE, Weng C, Skarda DE, Glissmeyer EW. Reduced computed tomography for evaluation of appendicitis in children after implementation of next-day surgery clinic follow-up. Pediatr Qual Saf. 2023;8:e641.

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260878

Magnetic resonance imaging radiomics for predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer

Emin Demirel[1]*, Huseyin Coskun[1]

SUMMARY

OBJECTIVE: The aim of this study was to develop and validate machine learning models integrating clinical and pre-treatment dynamic contrast-enhanced magnetic resonance imaging radiomic features for predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer.
METHODS: This retrospective study analyzed 299 breast cancer patients from the publicly available Duke Breast Cancer magnetic resonance imaging dataset who received neoadjuvant chemotherapy. From pre-treatment dynamic contrast-enhanced magnetic resonance imaging, 534 IBSI-compliant radiomic features were extracted. A five-method consensus feature selection pipeline retained 24 features. Notably, three prediction models were developed: clinical-only (Model A), radiomic-only (Model B), and combined clinical–radiomic (Model C). In total, seven machine learning algorithms were compared using 20 repeated five-fold stratified cross-validation with SMOTE applied within each training fold and assessed on a held-out test set (30%, n=90). Performance was evaluated using AUC with bootstrap confidence intervals, calibration (Brier score), and decision curve analysis.
RESULTS: The pathologic complete response rate was 20.4% (61/299). The combined model (Gradient Boosting) achieved the highest test AUC of 0.723 (95%CI 0.595–0.835), with sensitivity of 0.333 and specificity of 0.833. DeLong tests revealed no statistically significant differences between model pairs (all p>0.05). Decision curve analysis demonstrated the highest net benefit for the combined model across 10–50% threshold probabilities.
CONCLUSION: The combined clinical–radiomic model demonstrated modest discriminative ability for pathologic complete response prediction. The limited sensitivity suggests potential utility as a rule-out test to identify patients unlikely to achieve pathologic complete response. These hypothesis-generating findings require external validation in independent multi-center cohorts.
KEYWORDS: Breast cancer. Neoadjuvant chemotherapy. Pathologic complete response. Radiomics. Magnetic resonance imaging. Machine learning.

INTRODUCTION

Breast cancer is the most frequently diagnosed malignancy and leading cause of cancer-related mortality among women worldwide, with an estimated 2.3 million new cases annually[1]. Neoadjuvant chemotherapy (NAC) has become the standard of care for locally advanced and high-risk early stage breast cancers, particularly within triple-negative (TNBC) and HER2-positive subtypes[2,3], where pathologic complete response (pCR) rates have been reported to be highest[4]. NAC facilitates tumor downstaging, enabling conversion from mastectomy to breast-conserving surgery, and provides in vivo assessment of chemosensitivity.

pCR, defined as the absence of residual invasive carcinoma (ypT0/Tis ypN0) at surgical resection, serves as a well-established surrogate endpoint for long-term oncologic outcomes[5,6]. Patients achieving pCR demonstrate significantly improved event-free and overall survival, with the strongest association in aggressive subtypes[7]. However, pCR rates remain variable (20–40%), and many patients experience chemotherapy toxicity without therapeutic benefit[8]. This underscores the need for reliable pre-treatment biomarkers capable of identifying likely responders.

Conventional clinicopathologic parameters, including tumor grade, hormone receptor status, and Ki-67 index, have demonstrated limited predictive accuracy for pCR[9]. These markers are constrained by their reliance on small biopsy specimens that fail to represent intratumoral heterogeneity[10]. Imaging-based approaches, particularly MRI, have shown promise in this context. Dynamic contrast-enhanced MRI (DCE-MRI) is especially useful in women with dense breasts, where mammographic sensitivity is reduced[11].

Radiomics, the high-throughput extraction of quantitative features from medical images, has emerged as a promising approach for non-invasive tumor phenotyping[12,13]. DCE-MRI represents the reference-standard imaging modality in breast oncology, offering superior soft-tissue contrast and

[1]Afyonkarahisar University of Health Sciences, Faculty of Medicine, Department of Radiology – Afyonkarahisar, Turkey.
*Corresponding author: dremindemirel@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on June 03, 2026. Accepted on July 04, 2026.
Scientific Editor: José Maria Soares Júnior
Institution where work was performed: Department of Radiology, Faculty of Medicine, Afyonkarahisar University of Health Sciences, Afyonkarahisar, Turkey.

rich spatiotemporal data[14]. Quantitative radiomic features from pre-treatment DCE-MRI have demonstrated potential in capturing tumor pathophysiology and microarchitectural heterogeneity[15].

Despite this promise, methodological challenges have impeded clinical translation, including heterogeneous study designs, inconsistent feature extraction protocols, and overfitted models[16,17]. A comprehensive understanding of how molecular subtypes relate to imaging phenotypes is critical for interpreting radiomic findings[18]. The present study leverages the publicly available Duke Breast Cancer MRI dataset[19] to develop and validate prediction models integrating clinical and radiomic features for pCR prediction, in accordance with TRIPOD guidelines[20] and IBSI feature standardization[21].

METHODS

Study population and design

This retrospective study utilized the Duke Breast Cancer MRI dataset from The Cancer Imaging Archive[19]. From 922 female patients, 299 who received NAC with available pathologic response data were included. The primary outcome was pCR, defined as ypT0/Tis ypN0 (pCR: n=61, 20.4%; non-pCR: n=238, 79.6%). The dataset is publicly available and de-identified. Written informed consent was obtained from all participants at the time of original data collection. This research complied with the Declaration of Helsinki.

Magnetic resonance imaging acquisition and radiomic feature extraction

All examinations were performed using 1.5T or 3.0T MRI scanners with dedicated breast coils. DCE-MRI included pre-contrast T1-weighted and dynamic post-contrast acquisitions after gadolinium-based contrast agent administration (0.1 mmol/kg, 2 mL/s). Tumor regions of interest were delineated by experienced breast radiologists. A comprehensive set of 534 radiomic features was extracted using a custom Python pipeline built upon PyRadiomics (version 3.0.1, IBSI-compliant), encompassing tumor morphology, first-order statistics, gray-level co-occurrence matrix texture features, kinetic map features, and spatial autocorrelation metrics. Image preprocessing included z-score normalization with fixed bin width discretization.

Feature selection

Features underwent systematic quality control: removal of constant, near-zero variance, and highly correlated features (Pearson $|r|>0.95$), reducing the set from 534 to 286. A five-method consensus approach was implemented: (1) Mann-Whitney U test with FDR correction ($p<0.05$), (2) LASSO logistic regression (lambda.1se), (3) mutual information ranking (top 50), (4) recursive feature elimination with five-fold cross-validation, and (5) random forest importance (top 50). Features selected by at least two of the five methods were retained (n=24). When features were selected by exactly two methods (representing 11 of the 24 retained features), the direction of association with pCR was confirmed to be consistent between the two methods. This consensus threshold was set a priori to balance dimensionality reduction with retention of potentially informative features.

Model development

Notably, three prediction models were developed: Model A (clinical-only, 12 variables: age, ER status, PR status, HER2 status, tumor grade, HR subtype, TNBC status, tumor size, clinical stage, Ki-67 index, lymph node status, menopausal status), Model B (radiomic-only, 24 features), and Model C (combined, 36 features). In total, seven machine learning algorithms were evaluated: L2-regularized logistic regression (C=1.0), elastic net (alpha=0.5, l1_ratio=0.5), random forest (n_estimators=500), XGBoost (n_estimators=100, max_depth=4), LightGBM (n_estimators=100, num_leaves=31), SVM-RBF (C=1.0, gamma="scale"), and gradient boosting (n_estimators=100, max_depth=3). The dataset was partitioned into training (70%, n=209) and test (30%, n=90) sets with stratified sampling. SMOTE (k-neighbors=5)[20] was applied within each cross-validation training fold to prevent data leakage. Hyperparameter tuning used grid search with five-fold cross-validation.

Statistical analysis

Model performance was evaluated using AUC with 95%CIs from 1,000 bootstrap resamples, sensitivity, specificity, PPV, NPV, calibration, Brier score (ranging from 0 to 1, where lower values indicate better calibration), and decision curve analysis[21]. The Brier score represents the mean squared difference between predicted probabilities and observed outcomes; a score of 0 indicates perfect prediction, while 0.25 reflects a non-informative model for binary outcomes with approximately 50% prevalence. Model comparison used the DeLong test[22]. Internal validation employed 20 repeated five-fold stratified cross-validations. All analyses used Python 3.10 with scikit-learn (1.1.0), XGBoost (1.6.0), LightGBM (3.3.0), and SHAP (0.41.0). Statistical significance was set at two-sided $p<0.05$. The code underlying this analysis is available from the corresponding author upon reasonable request.

RESULTS

Study population and clinical characteristics

Among 299 patients, 61 (20.4%) achieved pCR. ER-negative status (p=0.001), PR-negative status (p<0.001), HER2-positive status (p=0.006), and triple-negative subtype (p<0.001) were significantly associated with higher pCR rates (Table 1). The molecular subtype distribution was: HR+/HER2- (46.5%), HR+/HER2+ (17.7%), HR-/HER2+ (8.7%), and TNBC (27.1%). TNBC accounted for 36.1% of pCR cases despite comprising only 27.1% of the cohort, consistent with the known association between TNBC and higher chemosensitivity.

Feature selection

The consensus feature selection identified 24 features selected by at least two methods. In total, three features were selected by all five methods: Tumor Grade (Mitotic), Change in Variance of Uptake (Tumor), and SER Map Max Probability (Tumor). Of the remaining features, 11 were selected by exactly two methods (with consistent direction of association confirmed), 7 by three methods, and 3 by four methods. The radiomic features span first-order statistics (n=8), texture/GLCM features (n=7), kinetic features (n=6), and morphological features (n=3).

Predictive model performance

The combined model (Model C, Gradient Boosting) achieved the highest test AUC of 0.723 (95%CI 0.595–0.835), with sensitivity of 0.333 and specificity of 0.833 (Table 2). The clinical-only model (Model A) achieved AUC of 0.700 (95%CI 0.581–0.819), while the radiomic-only model (Model B) achieved AUC of 0.615 (95%CI 0.468–0.748). Pairwise DeLong tests revealed no statistically significant differences between any model pair (all p>0.05). Internal validation via 20 repeated five-fold cross-validation yielded mean AUCs of 0.694 (SD: 0.091), 0.749 (SD: 0.080), and 0.764 (SD: 0.076) for Models A, B, and C, respectively.

The combined model achieved a Brier score of 0.190 (Model A: 0.209, Model B: 0.214), indicating acceptable calibration. Decision curve analysis demonstrated that the combined model provided the highest net benefit across threshold probabilities of 10–50%. Figure 1 presents the ROC curves for all three models.

DISCUSSION

In this exploratory prediction model study, the combined clinical–radiomic model achieved the numerically highest AUC of 0.723 for pCR prediction; however, the incremental

Table 1. Patient demographics and clinical characteristics.

Variable	All (n=299)	pCR (n=61)	Non-pCR (n=238)	p-value
Age, years, median (IQR)	48 (42–57)	47 (38–55)	49 (42–58)	0.062
ER positive, n (%)	184 (61.5)	26 (42.6)	158 (66.4)	0.001
PR positive, n (%)	151 (50.5)	14 (23.0)	137 (57.6)	<0.001
HER2 positive, n (%)	79 (26.4)	25 (41.0)	54 (22.7)	0.006
Grade 1/2/3	6/30/263	0/3/58	6/27/205	0.137
HR+/HER2-, n (%)	139 (46.5)	14 (23.0)	125 (52.5)	
HR+/HER2+, n (%)	53 (17.7)	15 (24.6)	38 (16.0)	
HR-/HER2+, n (%)	26 (8.7)	10 (16.4)	16 (6.7)	
TNBC, n (%)	81 (27.1)	22 (36.1)	59 (24.8)	<0.001

p-values from Mann-Whitney U test (continuous) and chi-square test (categorical). ER: estrogen receptor; PR: progesterone receptor; HER2: human epidermal growth factor receptor 2; HR: hormone receptor; TNBC: triple-negative breast cancer; IQR: interquartile range; pCR: pathologic complete response.

Table 2. Test set performance of best-performing models.

Model	Algorithm	AUC (95%CI)	Sensitivity	Specificity	PPV	Brier
Clinical (A)	Logistic Reg.	0.700 (0.581–0.819)	0.667	0.681	0.343	0.209
Radiomic (B)	Gradient Boost.	0.615 (0.468–0.748)	0.222	0.806	0.222	0.214
Combined (C)	Gradient Boost.	0.723 (0.595–0.835)	0.333	0.833	0.333	0.190

AUC: area under the receiver operating characteristic curve; CI: confidence interval; PPV: positive predictive value; Brier: Brier score. All metrics were calculated at the Youden J-statistic optimal threshold. Models: A=clinical-only (12 variables), B=radiomic-only (24 features), C=combined (36 features).

improvement over the clinical-only model was not statistically significant (DeLong p=0.780). This null finding is critical and must be interpreted with caution. The wide confidence interval reflects limited statistical power (n=90 test set, event rate ~20%), precluding definitive conclusions. Post-hoc consideration of sample size suggests that a formal power calculation would require substantially larger cohorts to detect moderate AUC improvements—our study was likely underpowered for this purpose. Prior studies reporting higher AUCs (0.65–0.92)[15,23] may reflect methodological differences, including smaller sample sizes, single-method feature selection, and less rigorous validation approaches. Our findings should be considered hypothesis-generating rather than confirmatory.

The biological interpretability of selected radiomic features provides important insights. Signal enhancement ratio (SER) and peak enhancement (PE) are directly related to tumor

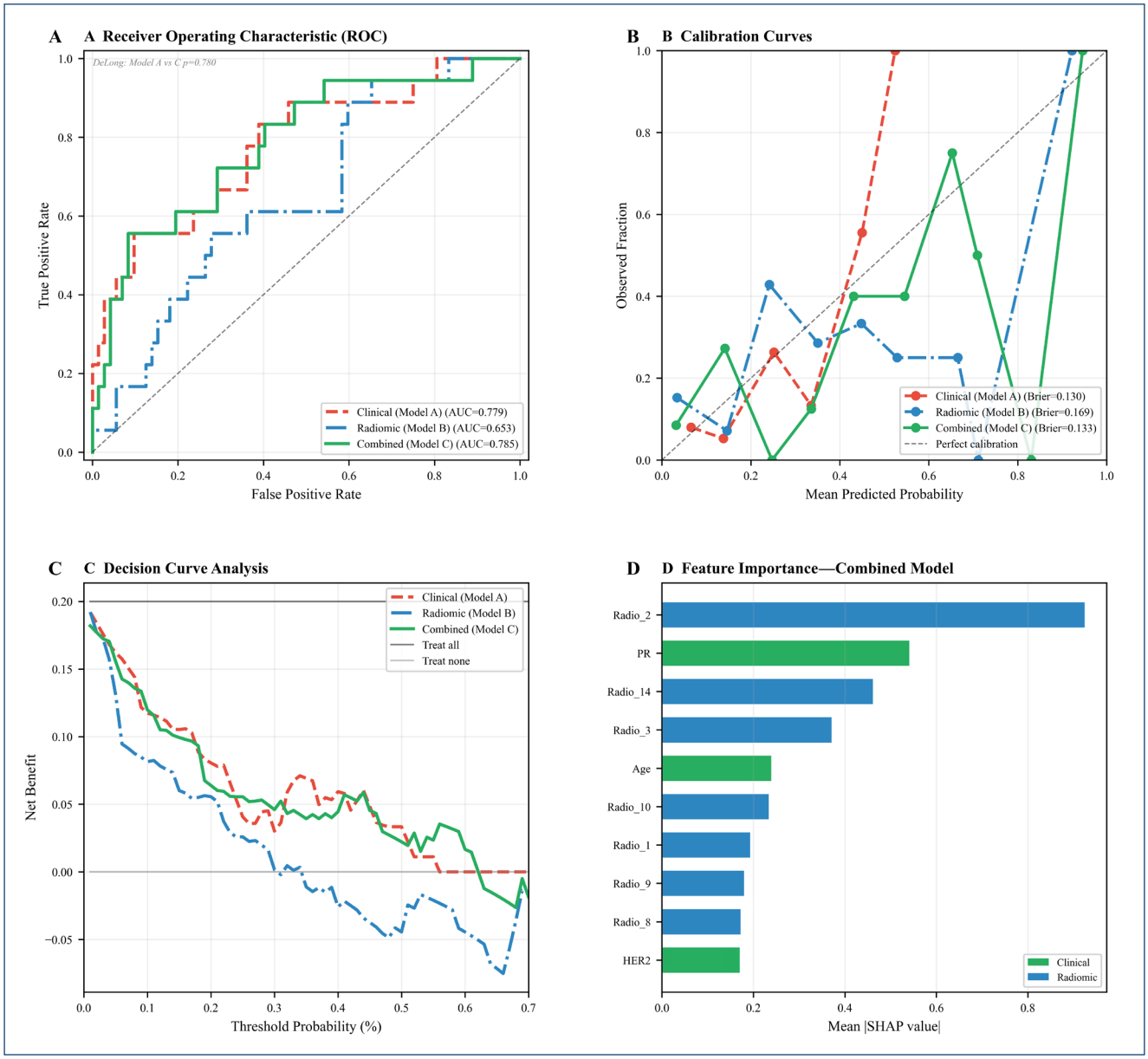


Figure 1. Four-panel multipanel figure. (A) Receiver operating characteristic (ROC) curves for the three prediction models on the held-out test set. The combined model (Model C, Gradient Boosting) achieved the highest AUC of 0.723 (95% CI 0.595–0.835). DeLong test results: Model A vs. B: p=0.321; Model A vs. C: p=0.780; Model B vs. C: p=0.153. (B) Calibration curves demonstrating agreement between predicted probabilities and observed outcomes. Brier scores: Model A=0.209, Model B=0.214, Model C=0.190. (C) Decision curve analysis showing net benefit across threshold probabilities. The combined model provided the highest net benefit across 10–50% threshold probabilities. (D) SHAP feature importance (top 10) for the combined model. Radiomic features in blue; clinical features in green.

angiogenesis and microvascular perfusion dynamics. High SER map max probability within the tumor reflects rapid contrast uptake and washout, indicative of abnormal vascular architecture driven by VEGF-mediated neovascularization. This high-perfusion profile may facilitate chemotherapeutic agent penetration, increasing pCR likelihood. Similarly, increased variance in tumor uptake reflects spatial heterogeneity, potentially harboring subclones with varying chemosensitivity.

A critical methodological consideration is potential confounding by molecular subtype. Our analysis shows that TNBC and HER2+ subtypes are strongly associated with pCR (Table 1), and these subtypes also have distinct imaging characteristics on DCE-MRI. It is possible that radiomic features partly capture subtype-specific imaging phenotypes rather than providing entirely independent predictive information. This is consistent with our observation that the clinical-only model (which includes subtype variables) performed comparably to the combined model. Future studies employing larger, subtype-stratified samples are needed to determine whether radiomic features provide independent predictive value beyond clinical subtype classification.

The clinical translation of our model must be carefully considered in light of its operating characteristics. With a sensitivity of 0.333 at the Youden-optimal threshold, the model would miss two-thirds of true pCR cases if used as a standalone diagnostic test. However, the high specificity (0.833) and reasonable net benefit on DCA suggest potential clinical utility as a rule-out test—identifying patients with very low probability of pCR who might benefit from alternative therapeutic strategies or closer monitoring during NAC. At the current performance level, we envision this model as a decision-support tool rather than a replacement for clinical judgment. A higher sensitivity threshold (e.g., targeting 0.80 sensitivity) could be clinically justified for specific scenarios but would substantially reduce specificity.

From a methodological perspective, our study distinguishes itself through strict adherence to TRIPOD guidelines and IBSI standards. The radiomic-only model exhibited a substantial gap between cross-validation AUC (0.749) and test AUC (0.615), with a drop of 0.134 AUC points, demonstrating persistent challenges of overfitting in high-dimensional radiomic studies despite multi-method feature selection. This gap, which reflects the broader challenge of radiomic feature generalizability, suggests that handcrafted radiomic features may capture dataset-specific noise rather than truly generalizable biological signals. Our five-method consensus approach reduced dimensionality from 534 to 24 features, representing a more robust alternative to single-method selection. Feature selection stability remains a concern: only three features were selected by all five methods, indicating that consensus-based approaches may benefit from bootstrap stability analysis to identify consistently selected features.

Our AUC of 0.723 falls within the moderate range of published studies (0.65–0.92)[15,16,23]. We chose conservative validation with honest performance reporting rather than optimistically biased results. The substantial gap between radiomic-only cross-validation and test performance (0.749 vs. 0.615) highlights a critical challenge in radiomics research: handcrafted features may generalize poorly compared to deep learning approaches that learn task-specific representations.

Several limitations warrant discussion. First, the absence of external validation is the most significant limitation—the Duke dataset is single-center, and no publicly available independent dataset with matched variables and NAC response data was identified for external testing. Second, exclusive use of pre-treatment imaging precludes capturing dynamic changes during NAC (delta-radiomics). Third, low sensitivity (0.333) limits standalone clinical utility. Fourth, inter-reader reproducibility was not independently assessed. Fifth, the non-significant DeLong tests indicate that the incremental value of radiomics over clinical variables was not statistically demonstrated. Sixth, our reliance on handcrafted radiomic features may not capture the full complexity of tumor biology that deep learning approaches might extract.

Recent work by Cristófalo et al.[4] provided contemporary data on pCR rates in TNBC, confirming that this subtype achieved the highest response rates and represented the subgroup where radiomic prediction models might be most clinically impactful. Mota et al.[11] highlighted the importance of advanced imaging in dense breasts, supporting our rationale for MRI-based radiomics. Rivas et al.[24] demonstrated the relevance of molecular subtype characterization in breast cancer detection, contextualizing our finding that subtype confounds radiomic-signal prediction. Future research should include external multi-center validation, delta-radiomics analysis, deep learning approaches, and multi-omics integration[24,25].

CONCLUSION

The combined clinical–radiomic model achieved the numerically highest predictive performance for pCR (test AUC=0.723, 95%CI 0.595–0.835). However, the incremental value of radiomic features over clinical variables alone was not statistically significant (DeLong p>0.05), and low sensitivity (0.333) limits standalone clinical deployment. Given its high specificity, the model may have utility as a rule-out test to identify patients unlikely to benefit from standard NAC who might be candidates for alternative

strategies or clinical trials. These findings are hypothesis-generating and require external validation in independent, multi-center cohorts before any clinical implementation.

ETHICS APPROVAL

The Duke Breast Cancer MRI dataset is publicly available and fully de-identified. Written informed consent was obtained from all participants at the time of original data collection. Additional ethics board review was not required. This research complied with the Declaration of Helsinki.

AUTHORS' CONTRIBUTIONS

ED: Conceptualization, Data curation, Formal Analysis, Methodology, Software, Visualization, Writing – original draft.
HC: Conceptualization, Methodology, Project administration, 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. 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. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet. 2014;384(9938):164-72. https://doi.org/10.1016/S0140-6736(13)62422-8
3. Kaufmann M, Minckwitz G, Mamounas EP, Cameron D, Carey LA, Cristofanilli M, et al. Recommendations from an international consensus conference on the current status and future of neoadjuvant systemic therapy in primary breast cancer. Ann Surg Oncol. 2012;19(5):1508-16. https://doi.org/10.1245/s10434-011-2108-2
4. Cristófalo MM, Maesaka JY, Pereira DA, Nóbrega GB, Reis YN, Soares Júnior JM, et al. Neoadjuvant chemotherapy in triple negative breast cancer: a systematic review of breast and node pathologic response. J Surg Oncol. 2026;133(6):712-9. https://doi.org/10.1002/jso.70213
5. Spring LM, Fell G, Arfe A, Sharma C, Greenup R, Reynolds KL, et al. Pathologic complete response after neoadjuvant chemotherapy and impact on breast cancer recurrence and survival: a comprehensive meta-analysis. Clin Cancer Res. 2020;26(12):2838-48. https://doi.org/10.1158/1078-0432.CCR-19-3492
6. Minckwitz G, Untch M, Blohmer JU, Costa SD, Eidtmann H, Fasching PA, et al. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J Clin Oncol. 2012;30(15):1796-804. https://doi.org/10.1200/JCO.2011.38.8595
7. Houssami N, Macaskill P, Minckwitz G, Marinovich ML, Mamounas E. Meta-analysis of the association of breast cancer subtype and pathologic complete response to neoadjuvant chemotherapy. Eur J Cancer. 2012;48(18):3342-54. https://doi.org/10.1016/j.ejca.2012.05.023
8. Haque W, Verma V, Butler EB, Teh BS. Response rates and pathologic complete response by breast cancer molecular subtype following neoadjuvant chemotherapy. Breast Cancer Res Treat. 2018;170(3):559-67. https://doi.org/10.1007/s10549-018-4801-3
9. Pilewskie M, Morrow M, Bao T, Eaton A, Patil S, King TA. Predictors of pathologic complete response after neoadjuvant chemotherapy in breast cancer. Ann Surg Oncol. 2016;23(2):468-74.
10. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer. 2012;12(5):323-34. https://doi.org/10.1038/nrc3261
11. Mota BS, Shimizu C, Reis YN, Gonçalves R, Soares Junior JM, Baracat EC, et al. Dense breasts and women's health: which screenings are essential? Clinics (Sao Paulo). 2025;80:100743. https://doi.org/10.1016/j.clinsp.2025.100743
12. Lambin P, Leijenaar RTH, Deistl TM, Peerlings J, Jong EEC, Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749-62. https://doi.org/10.1038/nrclinonc.2017.141
13. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278(2):563-77. https://doi.org/10.1148/radiol.2015151169
14. Mann RM, Cho N, Moy L. Breast MRI: state of the art. Radiology. 2019;292(3):520-36. https://doi.org/10.1148/radiol.2019182947
15. Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. Breast Cancer Res. 2017;19(1):57. https://doi.org/10.1186/s13058-017-0846-1
16. Liu Z, Li Z, Qu J, Zhang R, Zhou X, Li L, et al. Radiomics of multiparametric MRI for pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: a multicenter study. Clin Cancer Res. 2019;25(12):3538-47. https://doi.org/10.1158/1078-0432.CCR-18-3190
17. Saha A, Harowicz MR, Grimm LJ, Kim CE, Ghate SV, Walsh R, et al. A machine learning approach to radiogenomics of breast cancer: a study of 922 subjects and 529 DCE-MRI features. Br J Cancer. 2018;119(4):508-16. https://doi.org/10.1038/s41416-018-0185-8
18. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594. https://doi.org/10.1136/bmj.g7594
19. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping.

Radiology. 2020;295(2):328-38. https://doi.org/10.1148/radiol.2020191145

20. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321-57. https://doi.org/10.1613/jair.953
21. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565-74. https://doi.org/10.1177/0272989X06295361
22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837-45.
23. Granzier RWY, Nijnatten TJA,Woodruff HC, Smidt ML, Lobbes MBI. Exploring breast cancer response prediction to neoadjuvant systemic therapy using MRI-based radiomics: a systematic review. Eur J Radiol. 2019;121:108736. https://doi.org/10.1016/j.ejrad.2019.108736
24. Rivas FWS, Gonçalves R, Mota BS, Sorpreso ICE, Toporcov TN, Filassi JR, et al. Comprehensive diagnosis of advanced-stage breast cancer: exploring detection methods, molecular subtypes, and demographic influences — A cross-sectional study. Clinics (Sao Paulo). 2024;79:100510. https://doi.org/10.1016/j.clinsp.2024.100510
25. Calster B, McLernon DJ, Smeden M, Wynants L, Steyerberg EW, Topic Group "Evaluating Diagnostic Tests and Prediction Models" of the STRATOS Initiative. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. https://doi.org/10.1186/s12916-019-1466-7

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260890

A modified Rapid Emergency Medicine Score incorporating C-reactive protein for mortality prediction in septic shock

Serdar Özdemir[1]*, İbrahim Altunok[1], Abuzer Özkan[2], Hatice Şeyma Akça[3], Serkan Emre Eroğlu[1]

SUMMARY

OBJECTIVE: The aim of this study was to evaluate the prognostic performance of the Rapid Emergency Medicine Score incorporating C-reactive protein for predicting 30-day mortality in patients with septic shock and to compare its performance with the Rapid Emergency Medicine Score, National Early Warning Score, Modified Early Warning Score, quick Sequential Organ Failure Assessment, and Sequential Organ Failure Assessment.

METHODS: This retrospective observational study included 124 adult patients diagnosed with septic shock in a tertiary emergency department. Rapid Emergency Medicine Score incorporating C-reactive protein was calculated as Rapid Emergency Medicine Score +(C-reactive protein /50), using an a priori clinically determined C-reactive protein scaling factor.

RESULTS: The median age was 63 years, 52.4% of patients were female, and 76 patients (61.3%) died within 30 days. Rapid Emergency Medicine Score incorporating C-reactive protein had the highest numerical AUC (0.770), corresponding to moderate discrimination, followed by National Early Warning Score (0.755), Modified Early Warning Score (0.746), quick Sequential Organ Failure Assessment (0.688), Rapid Emergency Medicine Score (0.638), and Sequential Organ Failure Assessment (0.593). Rapid Emergency Medicine Score incorporating C-reactive protein had significantly higher AUCs than Rapid Emergency Medicine Score (ΔAUC=0.132, p<0.001) and Sequential Organ Failure Assessment (ΔAUC=0.177, p=0.049), whereas differences compared with National Early Warning Score, Modified Early Warning Score, and quick Sequential Organ Failure Assessment were not statistically significant. In the Firth-penalized logistic regression model, Rapid Emergency Medicine Score incorporating C-reactive protein remained associated with 30-day mortality (OR per 1-point increase, 1.22; 95%CI 1.11–1.36; p<0.001).

CONCLUSION: Rapid Emergency Medicine Score incorporating C-reactive protein showed moderate discrimination for 30-day mortality in this exploratory development cohort but did not demonstrate superior discrimination compared with National Early Warning Score, Modified Early Warning Score, or quick Sequential Organ Failure Assessment. Because Rapid Emergency Medicine Score incorporating C-reactive protein was developed and evaluated in the same cohort, these findings should be considered hypothesis-generating. Independent external validation is required before its clinical utility can be established.

KEYWORDS: Septic shock. Mortality. Risk assessment. C-reactive protein. Emergency service.

INTRODUCTION

Sepsis and septic shock are among the clinical conditions frequently encountered in emergency departments, characterized by systemic inflammatory response and multiple organ dysfunction and associated with high mortality rates. Accurate risk stratification at an early stage is critically important for preventing a fatal course[1-3]. Therefore, there is a growing need for reliable, rapid, and applicable scoring systems to predict prognosis in the management of patients with sepsis[4,5].

Various early warning systems developed for this purpose assist clinicians in patient monitoring and management. The National Early Warning Score (NEWS) allows dynamic monitoring of a patient's clinical condition based on vital signs. While effective in predicting in-hospital adverse outcomes, its specificity may be limited for rapidly progressing conditions like sepsis[6-8]. Similarly, the quick Sequential Organ Failure Assessment (SOFA) score has been recommended as a rapid screening tool for diagnosing sepsis; however, several studies have reported its inadequate sensitivity and insufficiency in supporting early intervention decisions[9,10].

Another early warning system is the Rapid Emergency Medicine Score (REMS), which includes parameters such as age, Glasgow Coma Scale (GCS) score, mean arterial pressure (MAP), heart rate, respiratory rate, and oxygen saturation[11]. REMS is widely used, particularly in emergency department settings[11], due to its practicality and predictive power in certain clinical situations[12,13]. However, because REMS is based exclusively on physiological parameters and

[1]Ümraniye Training and Research Hospital, Department of Emergency Medicine – Istanbul, Turkey.
[2]Taksim Training and Research Hospital, Department of Emergency Medicine – Istanbul, Turkey.
[3]Karamanoğlu Mehmet Bey University, Department of Emergency Medicine – Karaman, Turkey.
*Corresponding author: dr.serdar55@hotmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on June 26, 2026. Accepted on August 01, 2026.
Scientific Editor: José Maria Soares Júnior

does not incorporate biomarkers reflecting the inflammatory response, its prognostic performance may be limited in conditions characterized by profound systemic inflammation, such as septic shock[14,15]. We speculated that integrating biomarkers that directly represent the inflammatory response into these scoring systems could result in more precise models. Based on this premise, REMS-C, a new score created by integrating the C-reactive protein (CRP) level into REMS, constituted the focal point of this study. REMS-C was hypothesized to provide a more comprehensive assessment by considering both the patient's physiological stability and the level of inflammatory response.

This study aimed to determine the prognostic value of REMS-C in predicting 30-day mortality in patients presenting to the emergency department with septic shock and to evaluate its prognostic performance and compare it with REMS, NEWS, MEWS, and SOFA for predicting 30-day mortality.

METHODS

Study design and population

This retrospective observational study included adult patients diagnosed with septic shock and managed in the emergency department of a tertiary university hospital between June 1, 2021, and June 1, 2022.

Inclusion and exclusion criteria

Patients aged 18 years or older who met the Sepsis-3 criteria for septic shock at emergency department presentation were eligible for inclusion. Septic shock was defined as persistent hypotension requiring vasopressor support to maintain a MAP ≥65 mmHg and a serum lactate level >2 mmol/L despite adequate fluid resuscitation. Patients with incomplete clinical, laboratory, or outcome data were excluded from the study.

Data collection and score calculation

Demographic characteristics (age and sex), vital signs (heart rate, respiratory rate, systolic blood pressure, MAP, oxygen saturation, and body temperature), neurological status (GCS score and AVPU scale), laboratory parameters, and outcome data were retrospectively obtained from the hospital information management system and electronic medical records. Laboratory variables included pH, lactate, complete blood count parameters, inflammatory biomarkers, and routine biochemical measurements recorded at the time of emergency department admission.

REMS-C was developed as an exploratory modification of REMS by incorporating CRP as a marker of inflammatory burden. The CRP scaling factor was determined a priori on clinical grounds, before outcome analyses were performed. The upper reference limit for CRP at our institution is approximately 5 mg/L; therefore, 50 mg/L, corresponding to approximately 10 times this reference limit, was selected as a pragmatic and clinically interpretable increment. Accordingly, each 50-mg/L increase in CRP contributed one additional point to the modified score. This scaling factor was not derived from regression coefficients or selected by testing alternative weights to optimize predictive performance in the study cohort. REMS-C was calculated as follows: REMS-C=REMS+(CRP/50).

Statistical analysis

Statistical analyses were performed using Jamovi software (version 2.6.26.0; The Jamovi Project, Sydney, Australia). The distribution of continuous variables was assessed using the Shapiro-Wilk test. Continuous variables were presented as median and interquartile range (IQR, 25th–75th percentile), whereas categorical variables were presented as frequencies and percentages. Comparisons between survivor and non-survivor groups were performed using the Mann-Whitney U test for continuous variables and the chi-square test or Fisher's exact test, as appropriate, for categorical variables.

In an additional exploratory analysis, restricted cubic spline modeling with three knots at the 10th, 50th, and 90th percentiles of the CRP distribution was used to evaluate potential nonlinearity in the association between CRP and 30-day mortality. The spline model was compared with the corresponding linear model using a likelihood-ratio test.

Prognostic discrimination was evaluated using receiver operating characteristic (ROC) curve analysis, and differences between AUC values were assessed using the DeLong test.

To examine whether REMS-C was associated with 30-day mortality after adjustment for selected clinically relevant covariates while reducing small-sample bias, Firth-penalized logistic regression was performed. The model was specified parsimoniously and included REMS-C as the predictor of primary interest together with sex and serum lactate, which were selected on clinical grounds and are not components of REMS-C. Individual physiological variables incorporated within REMS-C were not entered simultaneously into the model to avoid predictor redundancy. Odds ratios (ORs) with 95%CIs were reported.

All statistical tests were two-tailed, and a p-value <0.05 was considered statistically significant.

RESULTS

A total of 124 patients diagnosed with septic shock were included in the study. Of these patients, 52.4% (n=65) were female, and the median age was 63 years (25th–75th percentile: 47–77). The 30-day mortality rate was 61.3% (n=76).

Univariate analyses revealed statistically significant differences between survivors and non-survivors across several clinical and laboratory parameters. Female sex, lower systolic and MAP, higher heart rate, higher respiratory rate, lower oxygen saturation, lower lymphocyte count, higher CRP levels, and higher REMS, NEWS, MEWS, and REMS-C scores were significantly associated with mortality. Detailed comparisons between the survivor and mortality groups are presented in Table 1.

Restricted cubic spline analysis showed no statistically significant evidence of nonlinearity in the association between CRP and 30-day mortality after adjustment for REMS (p for nonlinearity=0.164).

The prognostic performances of REMS-C, NEWS, MEWS, qSOFA, REMS, and SOFA for predicting 30-day mortality were evaluated using ROC curve analysis (Figure 1 and Table 2). REMS-C yielded an AUC of 0.770 (95%CI 0.681–0.858), while the AUCs were 0.755 for NEWS, 0.746 for MEWS, 0.688

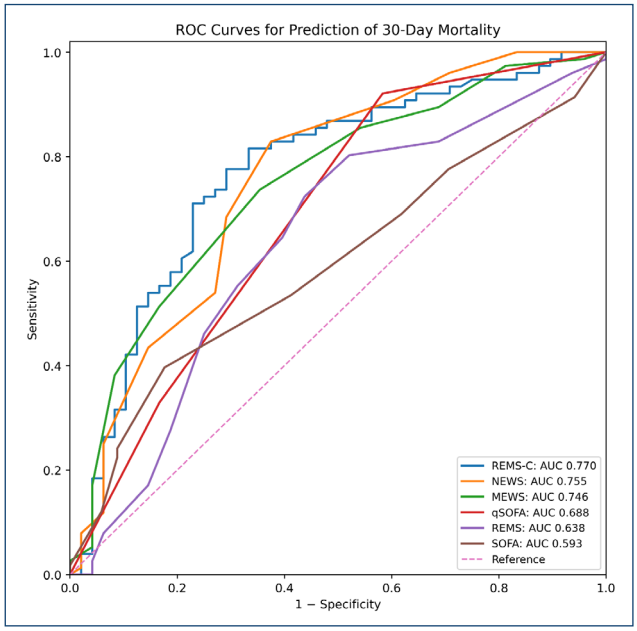


Figure 1. Receiver operating characteristic curves of Rapid Emergency Medicine Score incorporating C-reactive protein, National Early Warning Score, Modified Early Warning Score, quick Sequential Organ Failure Assessment, Rapid Emergency Medicine Score, and Sequential Organ Failure Assessment for predicting 30-day mortality in patients with septic shock.

Table 1. Descriptive characteristics of patients with septic shock and comparison between survivor and mortality groups.

Parameters	Total n=124	Survivor group n=48 (39%)	Mortality group n=76 (61%)	p-value
Age (years)	63 (47–77)	64 (57–76)	62 (39–78)	0.22
Gender				
Male	59 (47.6)	32 (66.7)	27 (35.5)	**<0.001**
Female	65 (52.4)	16 (33.3)	49 (64.5)	
Systolic blood pressure (mmHg)	77 (66–85)	81 (74–88)	70 (62–80)	**<0.001**
Diastolic blood pressure (mmHg)	45 (38–50)	48 (40–53)	44 (37–50)	**0.019**
Heart rate (beats per minute)	116 (102–130)	106 (90–121)	120 (110–133)	**<0.001**
Body temperature (°C)	37.00 (36.70–37.70)	36.90 (36.68–37.73)	37.00 (36.70–37.62)	0.78
Respiratory rate (breaths per minute)	24 (18–28)	19 (16–25)	25 (21–29)	**<0.001**
Mean arterial pressure (mmHg)	55 (48–62)	60 (51–64)	52 (45–60)	**<0.001**
Oxygen saturation (%)	89 (84–95)	92 (89–96)	86 (82–95)	**0.001**
Glasgow Coma Scale score	15 (12–15)	15 (14–15)	14 (10–15)	0.095
Rapid Emergency Medicine Score	9 (6–11)	7 (5–10)	10 (7–12)	**0.010**
National Early Warning Score	9 (7–11)	7 (5–10)	10 (8–11)	**<0.001**
Modified Early Warning Score	6 (4–8)	5 (3–6)	7 (5–8)	**<0.001**
pH	7.36 (7.27–7.42)	7.39 (7.29–7.44)	7.31 (7.24–7.40)	**0.017**
C-reactive protein (mg/L)	212 (116–296)	111 (68–183)	266 (193–341)	**<0.001**
Amylase (U/L)	39 (24–80)	50 (30–93)	33 (22–55)	**0.033**
SOFA score	8.00 (6.75–10.00)	8.00 (6.00–9.00)	9.00 (7.00–10.00)	0.14
REMS-C	13.2 (9.3–16.8)	9.5 (7.8–12.5)	15.7 (12.2–17.8)	**<0.001**

SOFA: Sequential Organ Failure Assessment; pCO_2: partial pressure of carbon dioxide; REMS-C: Rapid Emergency Medicine Score incorporating C-reactive protein. Bold values indicate statistically significant p-values (p<0.05).

Table 2. Prognostic performance of scoring systems for 30-day mortality.

Scoring system	AUC (95%CI)	Optimal cut-off	Sensitivity, %	Specificity, %	PPV, %	NPV, %
REMS-C	0.770 (0.681–0.858)	≥11.74	77.63	70.83	80.82	66.67
NEWS	0.755 (0.664–0.846)	≥8	82.89	62.50	77.78	69.77
MEWS	0.746 (0.657–0.834)	≥6	73.68	64.58	76.71	60.78
qSOFA	0.688 (0.597–0.779)	≥2	92.11	41.67	71.43	76.92
REMS	0.638 (0.534–0.742)	≥8	72.37	56.25	72.37	56.25
SOFA	0.593 (0.476–0.709)	≥10	39.66	82.35	79.31	44.44

AUC: area under the receiver operating characteristic curve; CI: confidence interval; MEWS: Modified Early Warning Score; NEWS: National Early Warning Score; NPV: negative predictive value; PPV: positive predictive value; qSOFA: quick Sequential Organ Failure Assessment; REMS: Rapid Emergency Medicine Score; REMS-C: Rapid Emergency Medicine Score–C-reactive protein; SOFA: Sequential Organ Failure Assessment.

(95%CI 0.597–0.779) for qSOFA, 0.638 for REMS, and 0.593 for SOFA. Pairwise DeLong analysis showed significantly higher AUCs for REMS-C than for REMS (ΔAUC=0.132, p<0.001) and SOFA (ΔAUC=0.177, p=0.049), whereas the differences compared with NEWS (p=0.754), MEWS (p=0.556), and qSOFA (ΔAUC=0.082, p=0.119) were not statistically significant.

Pairwise comparisons of AUC values using the DeLong method demonstrated that REMS-C significantly outperformed both REMS (ΔAUC=0.132, p<0.001) and SOFA (ΔAUC=0.177, p=0.049). However, no statistically significant differences were observed between REMS-C and NEWS (p=0.754) or between REMS-C and MEWS (p=0.556). Likewise, NEWS, MEWS, and SOFA did not significantly differ from one another in pairwise comparisons.

In the Firth-penalized logistic regression model including REMS-C, sex, and serum lactate, REMS-C remained associated with 30-day mortality (OR per 1-point increase, 1.22; 95%CI 1.11–1.36; p<0.001). Female sex was also associated with mortality (OR 3.00; 95%CI 1.30–6.91; p=0.010), whereas serum lactate was not independently associated with mortality (OR per 1 mmol/L, 1.04; 95%CI 0.89–1.21; p=0.604).

DISCUSSION

In this study, we evaluated the prognostic performance of REMS-C, a modified version of REMS incorporating CRP, for predicting 30-day mortality in patients presenting to the emergency department with septic shock. REMS-C yielded the highest numerical AUC (0.770) among the evaluated scores, corresponding to moderate discrimination. Although its AUC was significantly higher than those of REMS and SOFA, no statistically significant differences were observed compared with NEWS, MEWS, or qSOFA. Therefore, the present findings do not establish superiority of REMS-C over these widely used early warning scores. Rather, they provide exploratory evidence that incorporation of CRP into REMS warrants further investigation.

The operating characteristics of REMS-C also warrant cautious interpretation. At the identified cut-off of ≥11.74, sensitivity was 77.63%, specificity was 70.83%, PPV was 80.82%, and NPV was 66.67%. These values are insufficient to support REMS-C as a stand-alone rule-in or rule-out tool for individual clinical decisions. In addition, unlike NEWS and MEWS, REMS-C requires an available CRP measurement. Although CRP is commonly obtained during the evaluation of patients with suspected sepsis, the present study did not assess laboratory turnaround time, cost-effectiveness, workflow impact, or whether REMS-C-guided risk classification improves clinical outcomes. Therefore, its potential incremental clinical utility over purely physiological scores remains to be established prospectively.

In recent years, various early warning scoring systems have been developed and comparatively evaluated for the diagnosis and prognosis of sepsis in emergency departments. REMS has garnered increasing attention for this purpose. In a retrospective study evaluating data from 1,622 patients, REMS was reported to predict in-hospital mortality with significantly greater accuracy, with an AUC value of 0.62, compared to sepsis-specific scores, such as quick SOFA and Systemic Inflammatory Response Syndrome, and demonstrated comparable performance to NEWS[16]. Similarly, in another study involving 225 patients with sepsis in the emergency department, REMS yielded an AUC value of 88.6%, significantly higher than that of the Simple Clinical Score[17]. These findings underscore REMS as a valuable tool, especially due to its ease of use by frontline clinicians such as nurses. Nevertheless, in a prospective study involving patients receiving early goal-directed therapy for sepsis, the predictive power of REMS was found to be inferior to the Mortality in Emergency Department Sepsis score (AUC: 0.62 vs. 0.74), although it outperformed the CURB-65 score[18]. Previous studies have demonstrated the utility of

REMS in sepsis, with reported AUC values ranging from 0.62 to 0.89[19,20]. However, because REMS relies exclusively on physiological variables, it may not adequately reflect the inflammatory burden of septic shock. By incorporating CRP, REMS-C combines physiological and inflammatory information while maintaining bedside simplicity.

LIMITATIONS

This study has several limitations. First, its retrospective design precludes causal inference and may introduce selection bias. Second, the study was conducted at a single center and included a relatively limited number of patients, which may restrict the generalizability of the findings. Most importantly, REMS-C was developed and evaluated in the same cohort without an independent validation sample. Consequently, the reported discrimination represents apparent performance and may overestimate performance in new populations. Furthermore, the CRP scaling factor was determined a priori on pragmatic clinical grounds rather than derived from statistical modeling; therefore, the selected weighting should not be interpreted as an optimized or validated prognostic weight. The present findings should consequently be regarded as exploratory and hypothesis-generating. In addition, although REMS-C had a significantly higher AUC than REMS and SOFA, its discrimination was not significantly different from that of NEWS, MEWS, or qSOFA; therefore, its incremental clinical utility over existing scoring systems remains uncertain. Independent external validation in larger, prospective, multicenter cohorts with different demographic characteristics, comorbidity profiles, and infection sources is required before the clinical utility of REMS-C can be established.

CONCLUSION

In this exploratory proof-of-concept study, REMS-C showed moderate discrimination for 30-day mortality and had a higher numerical AUC than the other evaluated scores. Its discrimination was significantly higher than that of REMS and SOFA but not significantly different from NEWS, MEWS, or qSOFA. These findings do not establish incremental clinical utility or support REMS-C as a stand-alone bedside decision tool. Independent external validation and prospective assessment of clinical utility are required before its role in risk stratification can be determined.

AUTHORS' CONTRIBUTIONS

SÖ: Conceptualization, Data curation, Formal Analysis, Writing – original draft. **İA:** Data curation, Writing – review & editing. **AÖ:** Conceptualization, Data curation, Writing – review & editing. **HŞA:** Data curation, Writing – review & editing. **SEE:** Data curation, 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. Cecconi M, Evans L, Levy M, Rhodes A. Sepsis and septic shock. Lancet. 2018;392(10141):75-87. https://doi.org/10.1016/S0140-6736(18)30696-2
2. Long B, Gottlieb M. Emergency medicine updates: management of sepsis and septic shock. Am J Emerg Med. 2025;90:179-91. https://doi.org/10.1016/j.ajem.2025.01.054
3. Zhang N, Liu Y, Yang C, Li X. Review of the predictive value of biomarkers in sepsis mortality. Emerg Med Int. 2024;2024:2715606. https://doi.org/10.1155/2024/2715606
4. Chen YX, Li CS. Risk stratification and prognostic performance of the predisposition, infection, response, and organ dysfunction (PIRO) scoring system in septic patients in the emergency department: a cohort study. Crit Care. 2014;18(2):R74. https://doi.org/10.1186/cc13832
5. Liu B, Li D, Cheng Y, Yu J, Jia Y, Zhang Q, et al. Development and internal validation of a simple prognostic score for early sepsis risk stratification in the emergency department. BMJ Open. 2021;11(7):e046009. https://doi.org/10.1136/bmjopen-2020-046009
6. Abuzer Ö. Evaluation of short-term mortality prediction using initial lactate and NEWS+L at admission in COVID-19 patients. Disaster Med Public Health Prep. 2023;17:e333. https://doi.org/10.1017/dmp.2022.299
7. Usman OA, Usman AA, Ward MA. Comparison of SIRS, qSOFA, and NEWS for the early identification of sepsis in the emergency department. Am J Emerg Med. 2019;37(8):1490-7. https://doi.org/10.1016/j.ajem.2018.10.058
8. Oduncu AF, Kıyan GS, Yalçınlı S. Comparison of qSOFA, SIRS, and NEWS scoring systems for diagnosis, mortality, and morbidity of sepsis in emergency department. Am J Emerg Med. 2021;48:54-9. https://doi.org/10.1016/j.ajem.2021.04.006
9. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA. 2016;315(8):762-74. https://doi.org/10.1001/jama.2016.0288
10. Raith EP, Udy AA, Bailey M, McGloughlin S, MacIsaac C, Bellomo R, et al. Prognostic accuracy of the SOFA score, SIRS criteria, and qSOFA score for in-hospital mortality among adults with suspected infection admitted to the intensive care unit. JAMA. 2017;317(3):290-300. https://doi.org/10.1001/jama.2016.20328
11. Guan G, Lee CMY, Begg S, Crombie A, Mnatzaganian G. The use of early warning system scores in prehospital and emergency department settings to predict clinical deterioration: a systematic

review and meta-analysis. PLoS One. 2022;17(3):e0265559. https://doi.org/10.1371/journal.pone.0265559

12. İlhan B, Irmak MF, Tutar D, Eroğlu O, Deniz T. Physiological scoring systems for early identification of critically ill older adults in the emergency department: a prospective observational cohort study. BMC Emerg Med. 2025;25(1):218. https://doi.org/10.1186/s12873-025-01380-z
13. Olsson T, Terent A, Lind L. Rapid emergency medicine score: a new prognostic tool for in-hospital mortality in nonsurgical emergency department patients. J Intern Med. 2004;255(5):579-87. https://doi.org/10.1111/j.1365-2796.2004.01321.x
14. Özdemir S. Evaluation of the rapid emergency medicine score and rapid acute physiology score scoring systems in predicting short-term survival outcomes in geriatric septic shock patients in the emergency department. Am J Emerg Med. 2025;96:110-5. https://doi.org/10.1016/j.ajem.2025.06.034
15. Chang SH, Hsieh CH, Weng YM, Hsieh MS, Goh ZNL, Chen HY, et al. Performance assessment of the Mortality in Emergency Department Sepsis Score, Modified Early Warning Score, Rapid Emergency Medicine Score, and Rapid Acute Physiology Score in predicting survival outcomes of adult renal abscess patients in the emergency department. Biomed Res Int. 2018;2018:6983568. https://doi.org/10.1155/2018/6983568
16. Ruangsomboon O, Boonmee P, Limsuwat C, Chakorn T, Monsomboon A. The utility of the Rapid Emergency Medicine Score (REMS) compared with SIRS, qSOFA and NEWS for predicting in-hospital mortality among patients with suspicion of sepsis in an emergency department. BMC Emerg Med. 2021;21(1):2. https://doi.org/10.1186/s12873-020-00396-x
17. Chatchumni M, Maneesri S, Yongsiriwit K. Performance of the Simple Clinical Score (SCS) and the Rapid Emergency Medicine Score (REMS) to predict severity level and mortality rate among patients with sepsis in the emergency department. Australas Emerg Care. 2022;25(2):121-5. https://doi.org/10.1016/j.auec.2021.09.002
18. Crowe CA, Kulstad EB, Mistry CD, Kulstad CE. Comparison of severity of illness scoring systems in the prediction of hospital mortality in severe sepsis and septic shock. J Emerg Trauma Shock. 2010;3(4):342-7. https://doi.org/10.4103/0974-2700.70761
19. Jones AE, Trzeciak S, Kline JA. The Sequential Organ Failure Assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation. Crit Care Med. 2009;37(5):1649-54. https://doi.org/10.1097/CCM.0b013e31819def97
20. Jo S, Lee JB, Jin YH, Jeong TO, Yoon JC, Jun YK, et al. Modified early warning score with rapid lactate level in critically ill medical patients: the ViEWS-L score. Emerg Med J. 2013;30(2):123-9. https://doi.org/10.1136/emermed-2011-200760

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260891

Propofol versus remifentanil sedation for elective gastroscopy: a randomized controlled trial

Cansu Ofluoğlu[1]*, Doğa Meriç Yükselen[1], Arsen Güngör Ay[1], Ayşe Günler Gayretli[1], Öznur Demiroluk[1], Arzu Yıldırım Ar[1]

SUMMARY

OBJECTIVE: Propofol and remifentanil are widely used for elective gastroscopy sedation. The aim of this randomized controlled trial was to compare their safety profiles, patient-reported recovery quality, and endoscopist satisfaction.

METHODS: This single-center randomized controlled trial enrolled 88 adults (American Society of Anesthesiologists I–III) undergoing elective gastroscopy, allocated 1:1 to propofol (Group P, n=44) or remifentanil (Group R, n=44). Sedation depth was standardized using the Bispectral Index (target 60–80) and Ramsay Sedation Scale (target 2–3). The primary outcome was the Quality of Recovery-15 score at 30 min post-procedure; secondary outcomes included complication incidence, awakening time, and endoscopist satisfaction.

RESULTS: Quality of Recovery-15 scores were comparable (median 140 vs. 140; median difference 0, 95%CI -1.0 to 1.0; p=0.95). Total complication incidence was 0% in Group P versus 61.4% in Group R (p<0.001), comprising bradycardia (27.3%), postoperative nausea and vomiting (20.5%), and desaturation (13.6%). Awakening time was shorter in Group R (3 vs. 60 s; p<0.001). Endoscopist satisfaction was significantly higher in Group P (90.9 vs. 56.8% excellent scores; p=0.003).

CONCLUSION: Despite comparable patient-reported recovery quality, propofol was associated with substantially fewer perioperative adverse events (0 vs. 61.4%) and higher endoscopist satisfaction, supporting its preferential use in routine elective gastroscopy sedation.

Clinical Trial Registration Number: ClinicalTrials.gov, NCT07336628. Registered January 2, 2026.

KEYWORDS: Conscious sedation. Gastroscopy. Patient safety. Propofol. Postoperative period. Remifentanil.

INTRODUCTION

Gastroscopy is among the most frequently performed endoscopic procedures worldwide, and its use continues to increase[1]. Sedation is routinely employed to reduce anxiety, suppress the gag reflex, and improve procedural tolerability[2,3].

Propofol, a short-acting intravenous hypnotic, offers rapid onset and antiemetic properties. Remifentanil, an ultra-short-acting μ-opioid agonist with a context-sensitive half-time of approximately 3–4 min, is attractive for brief day-case procedures[4,5]. Each agent carries distinct adverse effects: propofol causes dose-dependent cardiorespiratory depression, while remifentanil may cause bradycardia, respiratory depression, and postoperative nausea and vomiting (PONV)[4,5].

Existing comparative studies have predominantly focused on colonoscopy or endoscopic retrograde cholangiopancreatography (ERCP), with gastroscopy-specific data limited and largely restricted to single safety endpoints[6,7]. The Quality of Recovery-15 (QoR-15) is a validated 15-item scale assessing five recovery dimensions; its application as a primary endpoint in gastroscopy sedation has not been previously reported[8,9].

This trial compared propofol and remifentanil sedation for elective diagnostic gastroscopy, using QoR-15 as the primary outcome. We hypothesized that propofol would demonstrate a more favorable safety profile with comparable patient-reported recovery quality.

METHODS

Study design and ethical approval

This prospective, single-center, randomized controlled trial was conducted between February and April 2026 at

[1]University of Health Sciences, Fatih Sultan Mehmet Training and Research Hospital, Department of Anesthesiology and Reanimation – Istanbul, Turkiye.

*Corresponding author: cansuakin.iu@gmail.com

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

Received on June 25, 2026. Accepted on August 15, 2026.

Scientific Editor: José Maria Soares Júnior

the Departments of Anesthesiology and Gastroenterology, University of Health Sciences, Fatih Sultan Mehmet Training and Research Hospital, Istanbul, Turkiye. Ethics committee approval was obtained (Code: 2025-KAEK-43; Decision No: 2025/20/1260; Date: November 4, 2025). Written informed consent was obtained from all participants. The study was conducted in accordance with the Declaration of Helsinki and prospectively registered at ClinicalTrials.gov (NCT07336628). This manuscript was prepared in accordance with Consolidated Standards of Reporting Trials (CONSORT) 2010 guidelines.

Study population

Adult patients (age ≥18 years, American Society of Anesthesiologists [ASA] I–III) scheduled for elective diagnostic gastroscopy under sedation were enrolled. Exclusion criteria were: patient refusal, emergency procedures, known allergy to propofol, remifentanil, or egg/soy products, and ASA IV or above.

Randomization and blinding

Eligible patients were randomized in a 1:1 ratio using a computer-generated block randomization sequence with block sizes of 4 and 8. The allocation sequence was prepared by an investigator who was not involved in patient recruitment, sedation, or outcome assessment and placed into sequentially numbered, sealed, opaque envelopes. Immediately before sedation, the attending anesthesiologist opened the next envelope in sequence to determine treatment allocation. Patients and the investigator assessing the QoR-15 questionnaire were blinded to treatment allocation. Complete endoscopist blinding was not feasible because of the distinct visual appearances of the study drugs.

Sedation protocols

Standard monitoring included electrocardiography, non-invasive blood pressure, pulse oximetry, and Bispectral Index (BIS) monitoring (BIS Vista, Medtronic). Supplemental oxygen was administered at 6 L/min via nasal cannula throughout the procedure. Group P received propofol (Propofol 1%, Fresenius Kabi) at an initial dose of 0.5–1 mg/kg, with additional doses administered as required to maintain a BIS value of 60–80 and a Ramsay Sedation Scale (RSS) score of 2–3. Group R received remifentanil (Ultiva, GlaxoSmithKline) as a 1 μg/kg intravenous loading dose followed by a continuous infusion of 0.025–0.1 μg/kg/min. The infusion rate was titrated throughout the procedure to maintain the same target BIS (60–80) and RSS (2–3) values and was discontinued immediately after completion of the gastroscopy. No midazolam premedication was administered.

Outcome measures

The primary outcome was the QoR-15 score measured 30 min after the procedure. The QoR-15 is a validated 15-item patient-reported measure of postoperative recovery[10], and quality of recovery is a recommended core endpoint for patient comfort in perioperative trials[11]. The validated Turkish version of the questionnaire was administered[12]. Because all participants underwent ambulatory day-case gastroscopy and were discharged on the same day, the questionnaire was administered before discharge to evaluate early postoperative recovery. Secondary outcomes included awakening time, perioperative complication incidence, procedure duration, time to procedure initiation, time to discharge, and endoscopist satisfaction assessed using a five-point Likert scale. Awakening time was defined as the interval, in seconds, from completion of the gastroscopy to the patient's first appropriate response to a verbal command. Complications were defined as hypotension (mean arterial pressure (MAP) <65 mmHg or >20% decrease from baseline), bradycardia (heart rate [HR] <50 bpm), desaturation (SpO_2 <90% persisting for more than 10 s), and PONV.

Statistical analysis

Sample size was calculated based on a QoR-15 minimal clinically important difference (MCID) of 8.0 points and standard deviation (SD) of 12 points (Cohen's d=0.67), yielding 37 patients per group (80% power, two-sided alpha 0.05); with a 20% dropout allowance, 44 per group were enrolled. Continuous variables are presented as median [interquartile range (IQR)] and compared using the Mann-Whitney U test; categorical variables were compared using chi-square or Fisher's exact test. Risk differences and risk ratios with 95%CI were calculated for binary outcomes. Analyses were performed using IBM SPSS v.25. A two-sided $p<0.05$ was considered significant.

RESULTS

Participant flow and baseline characteristics

Of 126 patients assessed, 38 were excluded (22 ASA IV or above, 6 emergency procedures, 8 declined consent, 2 other reasons). The remaining 88 were randomized; 44 to Group P and 44 to Group R. No dropouts occurred; all were included in the intention-to-treat analysis (Figure 1). Baseline

characteristics were comparable between groups (Table 1). The median total cumulative propofol dose was 160 (150–170) mg, whereas the median total cumulative remifentanil dose was 200 (200–280) µg.

Primary outcome: quality of recovery

BIS values were maintained within the target range (60–80), and RSS scores remained 2–3 throughout. No significant difference was observed in QoR-15 scores (p=0.95): median 140 (136–142) in Group P and 140 (134–144) in Group R, with a median difference of 0 points (95%CI -1.0 to 1.0), well below the MCID of 8.0 points (Table 2).

Perioperative complications and secondary outcomes

Desaturation occurred exclusively in Group R (13.6%, p=0.026). Bradycardia was recorded in 27.2% of Group R versus 0% in Group P (p<0.001). Total complication incidence was 0% in Group P versus 61.4% in Group R (p<0.001). PONV occurred in 20.5% of Group R patients versus 0% in Group P (p=0.003). Median awakening time was shorter in Group R (3 vs. 60 s; p<0.001). Procedure duration was longer in Group P (8 vs. 7 min; p=0.016). No significant difference in time to discharge was observed. Endoscopist satisfaction was significantly higher in Group P (90.9 vs. 56.8% excellent scores; p=0.003) (Table 2).

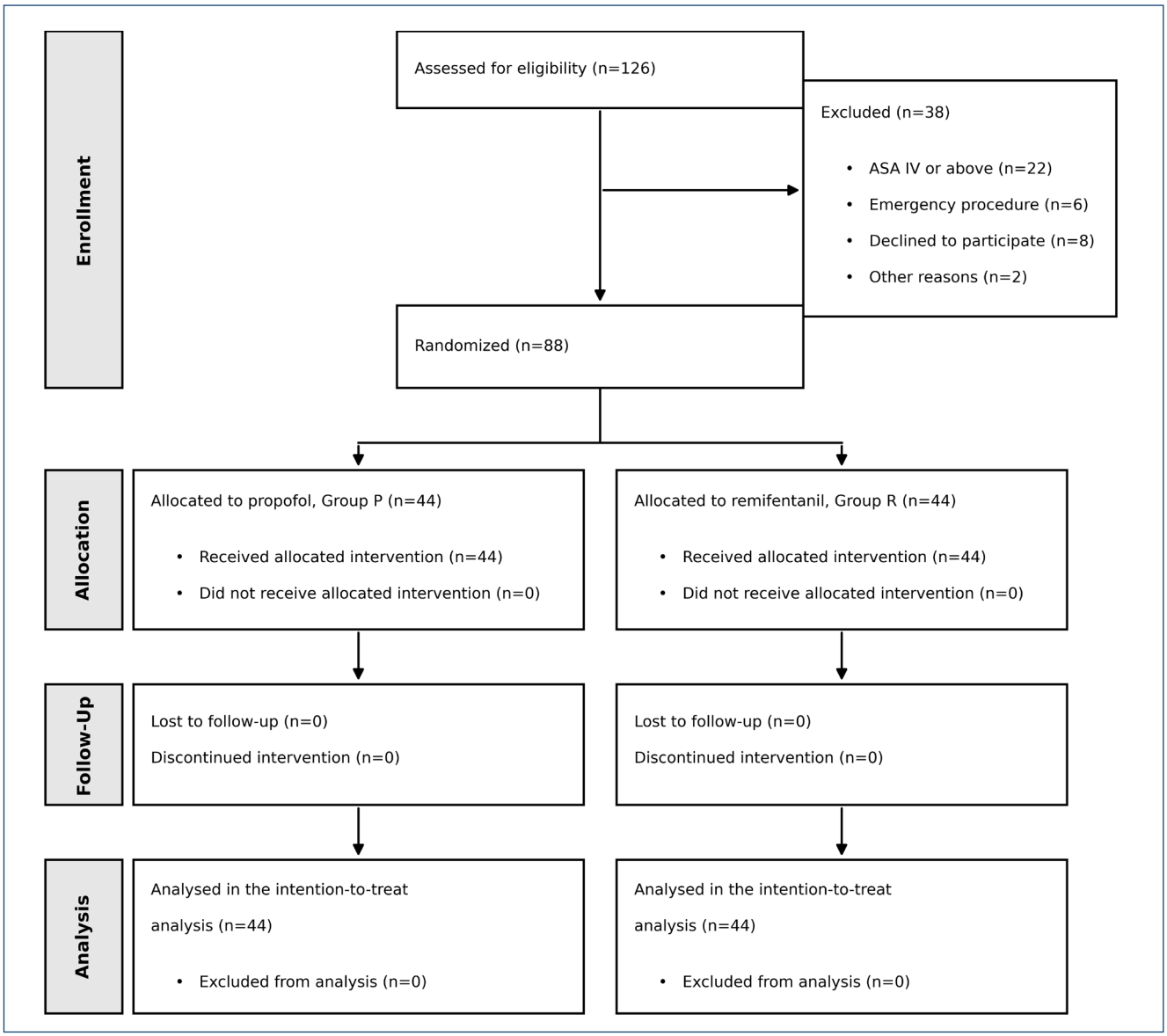


Figure 1. Consolidated Standards of Reporting Trials 2010 flow diagram illustrating participant screening, enrollment, allocation, and analysis. ASA: American Society of Anesthesiologists; ITT: intention-to-treat.

DISCUSSION

Both groups achieved median QoR-15 scores of 140, above the good recovery threshold of 122 points, with a between-group difference of 0 points (95%CI -1.0 to 1.0)—well below the MCID of 8.0 points[9,10]. The absence of a primary outcome difference is the principal finding; secondary safety differences are exploratory and should be interpreted accordingly. Previous comparative studies of sedative agents for gastrointestinal endoscopy have primarily focused on procedural success, cardiopulmonary adverse events, or patient satisfaction rather than patient-reported recovery quality. By using the QoR-15 as the primary endpoint, our study adds a multidimensional patient-centered assessment of early postoperative recovery. These findings suggest that comparable early recovery quality can be achieved despite substantial differences in perioperative safety profiles.

The complication profile in the remifentanil group reflects established pharmacology: μ-opioid receptor activation predisposes to bradycardia, while dose-dependent respiratory depression may precipitate desaturation[5]. All complications were brief and self-limiting, consistent with remifentanil's rapid offset kinetics. Nevertheless, a total complication

Table 1. Demographic characteristics and baseline variables.

Variable	Group P (n=44)	Group R (n=44)
Age, years—median (IQR)	51.5 (43.0–56.0)	52.0 (43.0–56.0)
BMI, kg/m²—median (IQR)	25.0 (24–28)	26.0 (26–32)
Sex—n (%)		
Female	28 (63.6%)	25 (56.8%)
Male	16 (36.4%)	19 (43.2%)
ASA score—n (%)		
ASA I	8 (18.2%)	6 (13.6%)
ASA II	28 (63.6%)	26 (59.1%)
ASA III	8 (18.2%)	12 (27.3%)
Total cumulative drug dose—median [IQR]	160 mg (150–170)	200 μg (200–280)

Data are presented as median (IQR) or n (%), as appropriate. Drug dose represents the total cumulative dose administered during the procedure. ASA: American Society of Anesthesiologists; BMI: body mass index; IQR: interquartile range.

Table 2. Perioperative outcome measures.

Outcome measure	Group P (n=44)	Group R (n=44)	p	RD or MD (95%CI)
QoR-15 score—median (IQR)	140 (136–142)	140 (134–144)	0.946	MD 0 (95%CI -1 to 1)
Procedure duration, min—median (IQR)	8 (8–10)	7 (6–8)	0.016*	–
Time to procedure initiation, sec—median (IQR)	60 (60–120)	47.5 (30–60)	0.016*	–
Awakening time, s—median [IQR]	60 (60–120)	3 (3–10)	<0.001	–
Bradycardia—n (%)	0 (0%)	12 (27.3%)	<0.001	RD -27.3% (-40.4% to -14.1%)
Desaturation—n (%)	0 (0%)	6 (13.6%)	0.026*	RD -13.6% (-23.8% to -3.5%)
PONV—n (%)	0 (0%)	9 (20.5%)	0.003*	RD -20.5% (-32.4% to -8.5%)
Hypotension—n (%)	0 (0%)	0 (0%)	–	–
Total complications—n (%)	**0 (0%)**	**27 (61.4%)**	**<0.001**	**RD -61.4% (-75.8% to -47.0%)**
Endoscopist satisfaction: Excellent (5)—n (%)	40 (90.9%)	25 (56.8%)	0.003	RD 34.1% (17.2–51.0%); RR 1.60 (1.22–2.10)
Endoscopist satisfaction: Good (4)—n (%)	4 (9.1%)	13 (29.5%)	–	–
Endoscopist satisfaction: Acceptable (3)—n (%)	0 (0%)	3 (6.8%)	–	–
Endoscopist satisfaction: Poor (2)—n (%)	0 (0%)	3 (6.8%)	–	–

CI: confidence interval; IQR: interquartile range; MD: median difference; PONV: postoperative nausea and vomiting; QoR-15: Quality of Recovery-15; RD: risk difference; RR: risk ratio. Mann-Whitney U test, chi-square test, or Fisher's exact test applied as appropriate. *p<0.05. Total complications indicates the number of patients who experienced at least one complication.

incidence of 61.4% represents a relevant safety concern in outpatient practice. The 20.5% PONV incidence is consistent with the emetogenic properties of opioids[13], while propofol's antiemetic effects via dopaminergic inhibition explain the complete absence of PONV in Group P. Our findings are consistent with previous studies reporting a higher incidence of opioid-related adverse events during endoscopic sedation with remifentanil, particularly bradycardia and PONV[14-16]. In contrast, propofol has consistently demonstrated a favorable safety profile and intrinsic antiemetic properties, supporting its widespread use for ambulatory gastrointestinal endoscopy[17].

The shorter awakening time with remifentanil (3 vs. 60 s; $p<0.001$) reflects its context-sensitive pharmacokinetics[5] but did not translate into reduced time to discharge, indicating that awakening speed alone does not determine discharge readiness. Rather, discharge after ambulatory gastroscopy depends on overall clinical recovery, including hemodynamic stability, adequate oxygenation, and the absence of treatment-related adverse events. Higher endoscopist satisfaction with propofol is consistent with the broader literature[17-20].

The present study has several strengths. Its prospective randomized controlled design minimized selection bias, while sedation depth was standardized using both BIS monitoring and the Ramsay Sedation Scale to ensure comparable sedation targets between groups. In addition, the QoR-15 questionnaire provided a validated patient-centered assessment of recovery, and complete follow-up of all randomized participants minimized the risk of attrition bias.

Limitations of this study include its single-center design and the inability to achieve complete endoscopist blinding. In addition, the QoR-15 questionnaire was administered 30 min after the procedure to assess early postoperative recovery before discharge because all participants underwent ambulatory day-case gastroscopy. Consequently, our findings reflect immediate recovery rather than the conventionally validated 24-h recovery profile. Finally, a ceiling effect was observed in both groups, which may have reduced the sensitivity of the QoR-15 to detect subtle differences in recovery quality.

CONCLUSION

Propofol and remifentanil provided comparable patient-reported recovery quality in elective gastroscopy. The 61.4% total complication incidence with remifentanil—versus zero with propofol—and the absence of a discharge time benefit support propofol as the preferred agent in routine outpatient gastroscopy. Remifentanil may remain appropriate where propofol is contraindicated, or rapid offset is specifically required. Larger multicenter trials are warranted.

ACKNOWLEDGMENTS

The authors thank the nursing and endoscopy staff of the Department of Gastroenterology for support during data collection.

ETHICS APPROVAL

This study was approved by the Institutional Ethics Committee of Kosuyolu Yuksek Ihtisas Egitim ve Arastirma Hastanesi, the review board affiliated with the study site (Approval Code: 2025-KAEK-43; Decision No: 2025/20/1260; Date: November 4, 2025).. The study was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants prior to enrollment.

AUTHORS' CONTRIBUTIONS

CO: Conceptualization, Data curation, Formal Analysis, Writing – original draft. **DMY:** Data curation. **AGA:** Data curation. **AGG:** Data curation. **ÖD:** Supervision, Writing – review & editing. **AYA:** Conceptualization, 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. Nurminen N, Järvinen T, Robinson E, Zhou N, Salo S, Räsänen J, et al. Upper gastrointestinal endoscopy procedure volume trends, perioperative mortality, and malpractice claims: population-based analysis. Endosc Int Open. 2024;12(3):E385-93. https://doi.org/10.1055/a-2265-8757
2. Sidhu R, Turnbull D, Haboubi H, Leeds JS, Healey C, Hebbar S, et al. British Society of Gastroenterology guidelines on sedation in gastrointestinal endoscopy. Gut. 2024;73(2):219-45. https://doi.org/10.1136/gutjnl-2023-330396
3. Correia C, Almeida N, Andrade R, Sant'Anna M, Macedo C, Perdigoto D, et al. Quality standards in upper gastrointestinal endoscopy: can deep sedation influence it? GE Port J Gastroenterol. 2024;31(2):101-9. https://doi.org/10.1159/000528977
4. Lv LL, Zhang MM. Up-to-date literature review and issues of sedation during digestive endoscopy. Wideochir Inne Tech

Maloinwazyjne. 2023;18(3):418-35. https://doi.org/10.5114/wiitm.2023.127854

5. Kapila A, Glass PS, Jacobs JR, Muir KT, Hermann DJ, Shiraishi M, et al. Measured context-sensitive half-times of remifentanil and alfentanil. Anesthesiology. 1995;83(5):968-75. https://doi.org/10.1097/00000542-199511000-00009
6. Motamed C, Servin F, Billard V. Adding low-dose propofol to limit anxiety during target-controlled infusion of remifentanil for gastrointestinal endoscopy: respiratory issues and safety recommendations. Medicina (Kaunas). 2022;58(9):1285. https://doi.org/10.3390/medicina58091285
7. Joy PJ, Blanshard HJ. Propofol-remifentanil patient-controlled sedation for endoscopic procedures: a prospective service audit. Can J Anaesth. 2023;70(11):1735-43. https://doi.org/10.1007/s12630-023-02593-1
8. Yahav-Shafir D, Orkin D, Zahavi G, Epstein I, Nadler M, Berkenstadt H. Patient-reported quality of recovery after sedation for endoscopy in the elderly. Gerontology. 2024;70(5):455-60. https://doi.org/10.1159/000536647
9. Myles PS, Shulman MA, Reilly J, Kasza J, Romero L. Measurement of quality of recovery after surgery using the 15-item quality of recovery scale: a systematic review and meta-analysis. Br J Anaesth. 2022;128(6):1029-39. https://doi.org/10.1016/j.bja.2022.03.009
10. Stark PA, Myles PS, Burke JA. Development and psychometric evaluation of a postoperative quality of recovery score: the QoR-15. Anesthesiology. 2013;118(6):1332-40. https://doi.org/10.1097/ALN.0b013e318289b84b
11. Myles PS, Boney O, Botti M, Cyna AM, Gan TJ, Jensen MP, et al. Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine (StEP) initiative: patient comfort. Br J Anaesth. 2018;120(4):705-11. https://doi.org/10.1016/j.bja.2017.12.037
12. Aslanlar E, Aslanlar DA, Doğanay C, Önal Ö, Sargin M, Çiçekci F, et al. The validity and reliability of the Turkish version of the quality of recovery-15 (QoR-15) questionnaire. Medicine (Baltimore). 2024;103(16):e37867. https://doi.org/10.1097/MD.0000000000037867
13. Gan TJ, Belani KG, Bergese S, Chung F, Diemunsch P, Habib AS, et al. Fourth consensus guidelines for the management of postoperative nausea and vomiting. Anesth Analg. 2020;131(2):411-48. https://doi.org/10.1213/ANE.0000000000004833
14. Komatsu R, Turan AM, Orhan-Sungur M, McGuire J, Radke OC, Apfel CC. Remifentanil for general anaesthesia: a systematic review. Anaesthesia. 2007;62(12):1266-80. https://doi.org/10.1111/j.1365-2044.2007.05221.x
15. Patel SS, Spencer CM. Remifentanil. Drugs. 1996;52(3):417-27. https://doi.org/10.2165/00003495-199652030-00009
16. Akcaboy ZN, Akcaboy EY, Albayrak D, Altinoren B, Dikmen B, Gogus N. Can remifentanil be a better choice than propofol for colonoscopy during monitored anesthesia care? Acta Anaesthesiol Scand. 2006;50(6):736-41. https://doi.org/10.1111/j.1399-6576.2006.01047.x
17. Padmanabhan A, Frangopoulos C, Shaffer LET. Patient satisfaction with propofol for outpatient colonoscopy: a prospective, randomized, double-blind study. Dis Colon Rectum. 2017;60(10):1102-8. https://doi.org/10.1097/DCR.0000000000000909
18. Lee CK, Lee SH, Chung IK, Lee TH, Park SH, Kim EO, et al. Balanced propofol sedation for therapeutic GI endoscopic procedures: a prospective, randomized study. Gastrointest Endosc. 2011;73(2):206-14. https://doi.org/10.1016/j.gie.2010.09.035
19. Wang D, Chen C, Chen J, Xu Y, Wang L, Zhu Z, et al. The use of propofol as a sedative agent in gastrointestinal endoscopy: a meta-analysis. PLoS One. 2013;8(1):e53311. https://doi.org/10.1371/journal.pone.0053311
20. Triantafyllou K, Sidhu R, Tham T, Tziatzios G, Guy C, Messmann H, et al. Sedation practices in gastrointestinal endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) survey. Endoscopy. 2024;56(12):964-74. https://doi.org/10.1055/a-2416-4866

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260904

Sodium glucose cotransporter 2 inhibitors may mask malignancy-associated fluorodeoxyglucose uptake in the thyroid on positron emission tomography/computed tomography

Recep Alanli[1]*, Murat Bulent Kucukay[1], Arzu Or Koca[2], Gulin Ucmak[3], Mutlu Dogan[4]

SUMMARY

OBJECTIVE: The aim of this study is to assess the effect of sodium glucose cotransporter 2 inhibitors on thyroid fluorodeoxyglucose uptake and maximum standardized uptake values on fluorine-18 fluorodeoxyglucose positron emission tomography/computed tomography, and to evaluate potential implications for the interpretation of thyroid incidentalomas.

METHODS: This retrospective study included 262 female breast cancer patients who underwent fluorine-18 fluorodeoxyglucose positron emission tomography/computed tomography imaging between July 2020 and July 2025. Patients were stratified into three groups: sodium glucose cotransporter 2 inhibitor users (n=55), diabetic patients using other antidiabetic agents (n=111), and non-diabetic controls (n=96). Thyroid fluorodeoxyglucose uptake was scored present or absent by a single blinded nuclear medicine physician; maximum standardized uptake value was measured only in avid glands (n=106).

RESULTS: Thyroid fluorodeoxyglucose uptake was less frequent in sodium glucose cotransporter 2 inhibitor users (p<0.001), and median maximum standardized uptake value among avid glands was lower (p=0.044). Sodium glucose cotransporter 2 inhibitor use remained independently associated with reduced uptake after adjustment for diabetes, glucose, glycated hemoglobin and thyroid stimulating hormone (p=0.008). In the receiver operating characteristics (ROC) analysis, an maximum standardized uptake value threshold of 6.37 yielded 75.0% specificity and 62.5% sensitivity for detecting malignant lesions. Four carcinomas, one in an sodium glucose cotransporter 2 inhibitor user, arose in non-avid glands.

CONCLUSION: Sodium glucose cotransporter 2 inhibitor use was independently associated with reduced thyroid fluorodeoxyglucose avidity and lower maximum standardized uptake value, unexplained by glycemic control or concurrent antidiabetic therapy. Clinicians should be aware of this pharmacological masking effect when evaluating thyroid incidentalomas in patients receiving sodium glucose cotransporter 2 inhibitors.

KEYWORDS: Fluorodeoxyglucose F18. Sodium-glucose transporter 2 inhibitors. Positron emission tomography computed tomography. Breast neoplasms. Thyroid nodule.

INTRODUCTION

18^{F}-fluorodeoxyglucose positron emission tomography/computed tomography (18^{F}-FDG PET/CT) is an established imaging method for evaluating patients with malignant and benign diseases. The technique exploits the higher glucose utilisation of tumour cells, mediated largely by the insulin-independent glucose transporter 1 (GLUT1), with 18^{F}-FDG metabolically trapped intracellularly after phosphorylation by hexokinase[1].

Glucose enters cells both by facilitated diffusion through the GLUT family and by secondary active transport through the sodium-glucose co-transporters (SGLT), which use the transmembrane sodium gradient. Several tumour types express SGLT, and SGLT2 inhibition is under investigation as an antineoplastic strategy[2-6]. Critically for imaging, however, 18^{F}-FDG is a substrate for the GLUT transporters and is not transported by SGLT[2]. Any effect of SGLT2 inhibition on 18^{F}-FDG biodistribution must therefore be indirect, arising from the systemic

[1]University of Health Sciences, Dr. Abdurrahman Yurtaslan Ankara Onkoloji Hospital, Gulhane Faculty of Medicine, Department of Internal Medicine – Ankara, Türkiye.
[2]University of Health Sciences, Dr. Abdurrahman Yurtaslan Ankara Onkoloji Hospital, Gulhane Faculty of Medicine, Department of Endocrinology – Ankara, Türkiye.
[3]University of Health Sciences, Dr. Abdurrahman Yurtaslan Ankara Onkoloji Hospital, Gulhane Faculty of Medicine, Department of Nuclear Medicine – Ankara, Türkiye.
[4]University of Health Sciences, Dr. Abdurrahman Yurtaslan Ankara Onkoloji Hospital, Gulhane Faculty of Medicine, Department of Medical Oncology – Ankara, Türkiye.
*Corresponding author: recepalanli@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on July 31, 2026. Accepted on August 08, 2026.
Scientific Editor: José Maria Soares Júnior

metabolic consequences of the drug rather than from competition at the transporter.

Some thyroid lesions discovered incidentally on 18^F-FDG PET/CT are malignant[7,8], and pooled data indicate that focal thyroid incidental uptake carries a substantial risk of malignancy[9,10]. The intensity of accumulation, expressed as the standardised uptake value (SUV), is a widely used semiquantitative parameter for clinical decision-making[11]. The diagnostic utility of SUVmax in thyroid malignancy nevertheless remains contested, since no universally accepted threshold reliably separates benign from malignant lesions[12-14].

Antidiabetic drugs are a recognised source of variability in tissue FDG uptake, and metformin and dipeptidyl peptidase-4 inhibitors have been reported to lower thyroid SUVmax[13,15]. SGLT2 inhibitors, now prescribed to a large proportion of patients with type 2 diabetes undergoing oncological staging, have been shown to reduce myocardial FDG uptake[16], but their effect on the thyroid gland has not been described, and current procedure guidelines contain no preparatory recommendation for them[17].

This study aimed to evaluate whether SGLT2 inhibitor use is associated with thyroid FDG uptake and SUVmax on 18^F-FDG PET/CT.

METHODS

This retrospective study included breast cancer patients who underwent 18^F-FDG PET/CT imaging between July 2020 and July 2025. This study was approved by the local ethics committee (Approval no: 2025-07/109, date: 24th July, 2025) and it was held in concordance with directives of the Helsinki Declaration and good clinical practices.

This study included 262 female breast cancer patients; aged over 18 years, who underwent 18^F-FDG PET/CT and thyroid ultrasonography. The patients under 18, patients who had total thyroidectomy operation, radioactive iodine treatment or thyroid malignancies before 18^F-FDG PET/CT procedure, and the patients who had other active malignancies than breast cancer, and the patients with history of Graves' disease, multinodular goiter or thyroiditis were excluded. The interval between thyroid ultrasonography and 18^F-FDG PET/CT more than one month was also an exclusion criterion.

Patients were stratified into three groups: SGLT2 inhibitor users (n=55), diabetic patients using other antidiabetic agents (n=111), and non-diabetic controls (n=96). To include a patient in the SGLT2 group, the medication should have been started at least six months prior to 18^F-FDG PET/CT.

Demographic characteristics, comorbidities, complete blood count parameters, glucose, thyroid hormones, glycated hemoglobin, and antithyroid antibodies were obtained from the hospital computer system. Thyroid fine needle aspiration biopsies were performed in accordance with the current guidelines[18].

Imaging was performed using a Biograph 6 Truepoint PET/CT system (Siemens Medical Solutions USA, Inc, Knoxville, USA). Prior to the 18^F-FDG PET/CT, patients followed the previously described preparation and treatment protocol[14]. Patients with completely normal physiological thyroid tissue (no significant visual uptake) did not have an SUVmax calculated or included in the metabolic numerical analysis.

Data were analyzed by Statistical Package for the Social Sciences (SPSS) software version 22.0 (IBM Corp., Armonk, NY, USA). Normally distributed continuous variables were reported as mean±standard deviation, while non-normally distributed data were expressed as median (interquartile range). Categorical variables were presented as numbers and percentages. Group comparisons were performed using the chi-square test, one-way analysis of variance (ANOVA), or the Kruskal-Wallis test, as appropriate for the data distribution. Correlations between variables were determined using Pearson or Spearman correlation analyses, also dependent on data distribution. To identify independent predictors of thyroid FDG uptake, variables with a $p<0.10$ in univariate analyses were referred to a multivariate logistic regression model. Statistical significance was defined as $p<0.05$. To address informative censoring arising from quantifying SUVmax only in avid glands, a two-part (Generalized Linear Model) model was employed.

RESULTS

In the SGLT2 inhibitor group, both the thyroid FDG uptake frequency on 18^F-FDG PET/CT ($p<0.001$) and median SUVmax values ($p=0.044$) were significantly lower compared with the other study groups. The correlation between SUVmax and SGLT2 inhibitor use was negatively significant ($r=-0.241$, $p=0.018$).

All participants were female, with a mean age of 55.80±9.65 years. Among SGLT2 inhibitor users, 58.2% received dapagliflozin 10 mg/day, while 36.4% received empagliflozin 10 mg/day and 5.4% received empagliflozin 25 mg/day. No association was observed between the specific type of SGLT2 inhibitor and thyroid FDG uptake ($r=0.078$, $p=0.298$). Smoking was more frequent in the control group ($p=0.035$). No significant differences were observed between SGLT2 inhibitor users and non-users regarding co-treatment with metformin (65.5 vs. 60.4%, $p=0.524$), Dipeptidyl Peptidase-4 (DPP-4)

inhibitors (23.6 vs. 23.4%, p=0.976), or exogenous insulin (45.5 vs. 43.2%, p=0.787).

Seventy-four patients (28.2%) were pre- or perimenopausal, and 188 (71.8%) were postmenopausal. Histopathological subtypes of breast cancer included invasive carcinoma in 228 patients (87.0%), lobular carcinoma in 22 (8.4%), metaplastic carcinoma in 6 (2.3%), and other subtypes in six patients (2.3%). Comparisons of demographic, laboratory, and imaging characteristics among the study groups are presented in Table 1. Thyroid autoimmune antibodies (anti-thyroid peroxidase and anti-thyroglobulin levels) were comparable among study groups (p=0.811 and p=0.434, respectively).

Table 1. The comparison of demographic, laboratory, and imaging data of the participants according to study groups.

Variable	Group 1 SGLT2i (n=55)	Group 2 T2DM (n=111)	Group 3 control (n=96)	p-value*
Demographic/clinical features				
Age (year)	57.0±9.39	56.59±9.69	54.21±9.64	0.122
Body mass index (kg/m^2)	29.14±3.10	28.73±4.59	23.5±3.96	0.202
Breast cancer				
Stage 1	3 (5.5%)	11 (9.9%)	2 (2.1%)	0.008§
Stage 2	20 (36.4%)	27 (24.3%)	48 (50.0%)	
Stage 3	24 (43.6%)	56 (50.5%)	36 (37.5%)	
Stage 4	8 (14.5%)	17 (15.3%)	10 (10.4%)	
Laboratory/imaging features				
Neutrophil (×10^9/L)	4.94±2.59	4.39±2.01	3.97±1.61	0.018‖
Lymphocyte (×10^9/L)	2.04±0.92	1.99±0.90	1.82±0.68	0.221
Hemoglobin (g/L)	125.5±15.6	126±15.3	128.9±13.0	0.234
Glucose (mmol/L)	7.28±1.46	7.17±1.53	5.64±1.12	0.001¶
Glycated hemoglobin (%)	8.0 (7–9.1)	6.9 (6.4–8.5)	5.6 (5.4–6)	<0.001**
TSH (mIU/L)	1.5 (1–2.9)	1.7 (1–3)	1.6 (0.9–2.5)	0.417
Free T4 (pmol/L)	15.70±4.38	15.58±4.76	15.70±5.28	0.993
Free T3 (pmol/L)	3.89±0.85	5.02±1.78	4.36±1.11	0.120
C-reactive protein (nmol/L)	60.0 (24.8–90.5)	57.1 (31.4–119.0)	44.8 (20.0–91.4)	0.196
Thyroid nodule (length, mm)	12.5 (8.8–22.5)	15 (11–23)	13 (10–18)	0.266
Pattern of thyroid nodule				
Hypoechoic	7 (43.8%)	24 (43.7%)	26 (44.8%)	0.601
Isoechoic	9 (56.2%)	29 (52.7%)	32 (55.2%)	
Hyperechoic	0	2 (3.6%)	0	
Presence of uptake in thyroid tissue	10 (18.2%)	48 (43.2%)	48 (50.0%)	<0.001†
Pattern of uptake in thyroid tissue				
Diffuse	4 (40.0%)	11 (23.4%)	11 (22.9%)	0.224
Heterogenous	3 (30.0%)	25 (53.2%)	17 (35.4%)	
Focal	3 (30.0%)	11 (23.4%)	20 (41.7%)	
SUVmax (only in cases with measurable uptake, n=106)	3.9 (2.2–4.3, n=10)	5.1 (3.7–6.7, n=48)	5.7 (3.8–7.0, n=48)	0.044‡
Thyroid malignancy	3 (5.5%)	9 (8.1%)	6 (6.3%)	0.449

*Data presented as mean±std. deviation, median (interquartile range) and number (percentages) were analyzed by One-way ANOVA test, Kruskal-Wallis and Chi-square tests respectively, §Stage 2 breast cancer, group 2–3, ‖Group 1–3, p=0.015, ¶Group 1–3 and 2–3 p<0.001, **Group 1–3 and 2–3 p<0.001, group 1–2 p=0.017, †Group 1–2 and 1–3, ‡Group 1–3: p=0.038, SGTL2i: sodium glucose co-transporter 2 inhibitors; T2DM: type 2 diabetes mellitus; TSH: thyroid stimulating hormone; SUVmax: maximum standardized uptake value.

Histological subtypes of breast cancer showed no significant association with the presence of thyroid FDG uptake (p=0.971) or median thyroid SUVmax values (p=0.381). Cancer stage was not associated with thyroid FDG uptake frequency (p=0.643) or thyroid levels (p=0.621), confirming that primary tumor stage did not confound thyroid metabolic measurements.

SUVmax showed a positive correlation with the presence of thyroid malignancy (r=0.303 [95%CI 0.033–0.533], p=0.029) and a negative correlation with SGLT2 inhibitor use (r=-0.241 [95%CI -0.422 to 0.042], p=0.018). In multivariate logistic regression analysis, SGLT2 inhibitor use remained an independent factor associated with reduced significant thyroid FDG uptake on 18^{F}-FDG PET/CT (Table 2). In the two-part hurdle model, multivariable Gamma regression demonstrated that SGLT2 inhibitor use independently predicted lower conditional SUVmax intensity (p=0.003) and adjusted estimated marginal mean SUVmax was 44% lower (3.07 [95%CI 2.20–4.28] in users vs. 5.43 [95%CI 4.67–6.30] in non-users).

Fine-needle aspiration biopsy was performed in 89 patients (33.9%) with suspicious thyroid findings on PET/CT and/or ultrasonography. Biopsies were conducted in 9 (16.4%) patients in the SGLT2i group, 41 (36.9%) in the diabetes mellitus (DM) group, and 39 (40.6%) in controls. Overall, 18 patients (6.9%) were diagnosed with thyroid carcinoma: 15 papillary, 2 oncocytic, and 1 follicular. Thyroid cancer detection rates among patients with versus without PET uptake were 66.7% (n=2) versus 16.7% (n=1) in the SGLT2i group (p=0.226), 26.9% (n=7) versus 13.3% (n=2) in the DM group (p=0.311), and 21.7% (n=5) versus 6.2% (n=1) in controls (p=0.187).

ROC curve analysis of standardized uptake values for diagnosis of thyroid malignancy was shown in Figure 1. SUVmax=6.37, yielding a sensitivity of 62.5% and a specificity of 75.0% for diagnosis of thyroid malignancy. Exploratory, two alternative operating points were examined: 4.37 favored sensitivity (81.3%, specificity 52.8%) and 7.24 favored specificity (88.9%, sensitivity 43.8%). No threshold achieved acceptable sensitivity and specificity simultaneously.

DISCUSSION

According to the results of this study, both the SUVmax and the prevalence of increased thyroidal uptake on 18^{F}-FDG PET/CT were significantly lower in the SGLT2 inhibitor group. A negative correlation was observed between SUVmax and SGLT2 inhibitor use. These findings suggest that SGLT2 inhibitor use may reduce FDG uptake in thyroid tissue independently

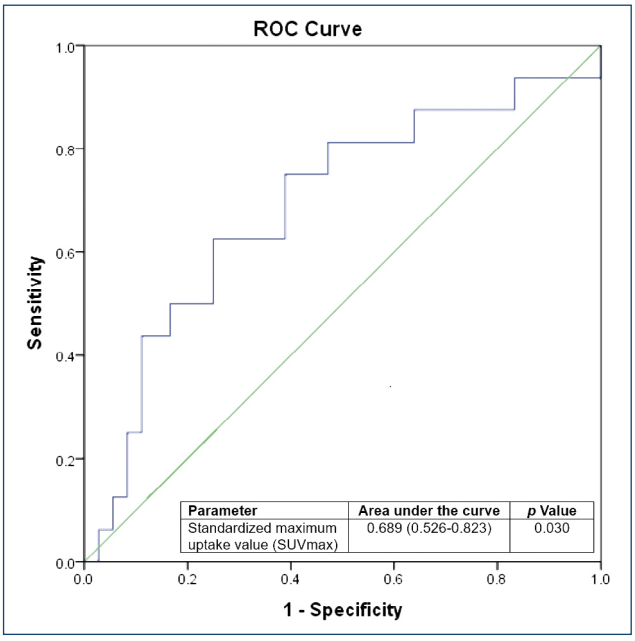


Figure 1. Receiver operator characteristics curve analysis of standardized uptake value for diagnosis of thyroid malignancy.

Table 2. Multivariable logistic regression analysis of independent predictors for thyroid fluorodeoxyglucose uptake*.

Parameters	Univariate logistic regression		Multivariate logistic regression	
	OR (95%CI)	p-value	OR (95%CI)	p-value
SGLT2 inhibitor	0.257 (0.123–0.537)	<0.001	0.292 (0.118–0.725)	0.008
Age	1.018 (0.992–1.045)	0.179		
Body mass index	1.001 (0.850–1.179)	0.990		
Smoking status	0.977 (0.508–1.882)	0.945		
Presence of DM	0.537 (0.322–0.896)	0.017	0.568 (0.140–2.310)	0.430
Glucose	0.987 (0.979–0.997)	0.007	0.998 (0.990–1.006)	0.617
Glycated hemoglobin levels	0.834 (0.667–1.042)	0.110	0.890 (0.689–1.149)	0.370
Thyroid-stimulating hormone	1.113 (0.986–1.256)	0.083	1.064 (0.892–1.269)	0.490

*Dependent variable: Thyroid FDG uptake. OR: odds ratio; CI: confidence interval; SGLT2: sodium glucose cotransporter 2; DM: diabetes mellitus.

of glycemic control, concurrent antidiabetic agents and other potential confounding factors.

The observed reduction in thyroid FDG uptake likely results from systemic metabolic reprogramming rather than direct transporter competition. Since FDG is a GLUT substrate, SGLT2 inhibition cannot compete directly at the thyrocyte membrane. Instead, the effect is attributed to drug-induced shifts toward glucosuria and fatty acid oxidation, consistent with previously reported[16] reductions in myocardial FDG avidity. Clinically, this pharmacological masking means that low PET/CT intensity in SGLT2 inhibitor users must not provide false reassurance. Physicians should incorporate medication reconciliation into reporting and maintain a low threshold for dedicated neck ultrasonography whenever structural nodularity is suspected, irrespective of reassuring SUVmax values.

Metformin and DPP-4 inhibitors were found to lower thyroid SUVmax in diabetic patients with thyroid cancer[13,15]. However, another study reported comparable SUVmax values between patients with breast cancer with and without DM[19]. Current EANM procedure guidelines do not provide specific recommendations regarding SGLT2 inhibitors[17]. In our cohort, thyroid avidity was significantly lower in diabetic patients receiving SGLT2 inhibitors compared to both non-SGLT2-treated diabetic patients and non-diabetic controls, despite similar exposure to metformin, DPP-4 inhibitors, and insulin. This balanced baseline supports a drug-class-specific effect rather than an artifact of diabetes or co-therapies. Furthermore, the SGLT2 group exhibited the lowest uptake despite poorer glycemic control, suggesting the reduction results from systemic metabolic reprogramming rather than direct substrate competition. SGLT2 inhibitors are characteristically added to regimens that have failed to reach target, reflecting the inadequate control; so treated patients are selected for higher glycated hemoglobin. Clinically, this pharmacological masking means low PET/CT intensity in SGLT2 inhibitor users should not provide false reassurance; sonographic assessment of suspicious nodules should not be deferred regardless of SUVmax values.

Incidental thyroid hypermetabolism on PET/CT often prompts further work-up[20], as many incidentalomas are malignant[21,22]. The prevalence of focal thyroid incidental uptake at about 2% of scans with a malignancy risk of roughly one in three[9,10]. Our study's SUVmax threshold for malignancy (6.37) aligns with external data[22] (~6.25). Crucially, while current guidelines prioritize sonographic morphology over PET intensity, they do not account for medication-induced metabolic attenuation. SGLT2 inhibitors reduce thyroid FDG uptake, meaning a low or absent PET signal in these patients has reduced negative predictive value. We therefore recommend documenting SGLT2 inhibitor use in PET reports and ensuring that low SUVmax values do not deter sonographic assessment of structurally suspicious nodules.

This study has several limitations. The sample size was determined by the availability of eligible patients during the study period, and was not calculated a priori. The retrospective design precludes causal inference and cannot exclude confounding by indication. SUVmax was quantified only in avid glands, so that comparison is conditional on avidity in a non-random way, with small subgroups. Histopathology was available in 89 patients with only 18 carcinomas, leaving subgroup and threshold analyses underpowered. European Thyroid Imaging and Reporting Data System (EU-TIRADS) categories, diabetes duration, cumulative exposure, adherence and breast cancer treatment details were not reliably recoverable. Also, all PET/CT interpretations were performed by a single nuclear medicine physician so the visual assessment protocol was not validated and interobserver variability was not available.

CONCLUSION

This study demonstrates that SGLT2 inhibitor therapy is independently associated with significantly lower FDG uptake and reduced SUVmax values in thyroid tissue on 18^{F}-FDG PET/CT, irrespective of glycemic control and other antidiabetic medications or serum thyroid stimulating hormone. Clinically, while elevated thyroid FDG uptake (SUVmax$\geqslant$6.37) retains high specificity for thyroid malignancy, low or absent uptake in patients receiving SGLT2 inhibitors cannot safely rule out carcinoma. This potential metabolic masking effect highlights that FDG PET/CT intensity alone should not dictate diagnostic management. In patients taking SGLT2 inhibitors, thyroid biopsy decisions must be guided primarily by sonographic risk stratification and nodule morphology.

ETHICAL APPROVAL

This study was approved by the University of Health Sciences, Istanbul, Gulhane Faculty of Medicine, Dr. A. Y. Ankara Oncology Hospital Ethics Committee for Observational Studies with decision number 2025-07/109 (24 July 2025).

AUTHORS' CONTRIBUTIONS

RA: Conceptualization, Data curation, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing. **MBK:** Conceptualization, Data curation, Methodology, Software, Supervision, Writing original draft, Writing – review

& editing,. **AOK:** Data curation, Methodology, Writing – review & editing. **GU:** Data curation, Methodology, Writing – review & editing. **MD:** Conceptualization, Methodology, 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. Prabhu M, Nisamudeen F, Vikas H. [18F]FDG PET/CT in benign and malignant diseases involving the biliary tract. Indian J Nucl Med. 2024;39:185-90. https://doi.org/10.4103/ijnm.ijnm_5_23
2. Mahato A, Jain A, Prakash VS, Nair R, Joshi R, Paliwal D, et al. Use of oral empagliflozin to obtain optimal blood sugar levels for conducting 18 F-FDG PET-CT in patients with hyperglycemia-a pilot study. World J Nucl Med. 2023;22:191-5. https://doi.org/10.1055/s-0043-1771283
3. Klimek K, Chen X, Sasaki T, Groener D, Werner RA, Higuchi T. PET imaging of sodium-glucose cotransporters (SGLTs): unveiling metabolic dynamics in diabetes and oncology. Mol Metab. 2024;90:102055. https://doi.org/10.1016/j.molmet.2024.102055
4. Wang Y, Yang L, Mao L, Zhang L, Zhu Y, Xu Y, et al. SGLT2 inhibition restrains thyroid cancer growth via G1/S phase transition arrest and apoptosis mediated by DNA damage response signaling pathways. Cancer Cell Int. 2022;22:74. https://doi.org/10.1186/s12935-022-02496-z
5. Basak D, Gamez D, Deb S. SGLT2 inhibitors as potential anticancer agents. Biomedicines. 2023;11:1867. https://doi.org/10.3390/biomedicines11071867
6. Tsunokake S, Iwabuchi E, Miki Y, Kanai A, Onodera Y, Sasano H, et al. SGLT1 as an adverse prognostic factor in invasive ductal carcinoma of the breast. Breast Cancer Res Treat. 2023;201:499-513. https://doi.org/10.1007/s10549-023-07024-9
7. Du J, Gu J, Deng J, Kong L, Guo Y, Jin C, et al. The expression and survival significance of sodium glucose transporters in pancreatic cancer. BMC Cancer. 2022;22:116. https://doi.org/10.1186/s12885-021-09060-4
8. Leijer JF, Metman MJH, Hoorn A, Brouwers AH, Kruijff S, Hemel BM, et al. Focal thyroid incidentalomas on 18F-FDG PET/CT: a systematic review and meta-analysis on prevalence, risk of malignancy and inconclusive fine needle aspiration. Front Endocrinol (Lausanne). 2021;12:723394. https://doi.org/10.3389/fendo.2021.723394
9. Treglia G, Bertagna F, Sadeghi R, Verburg FA, Ceriani L, Giovanella L. Focal thyroid incidental uptake detected by 18F-fluorodeoxyglucose positron emission tomography. Meta-analysis on prevalence and malignancy risk. Nuklearmedizin. 2013;52:130-6. https://doi.org/10.3413/Nukmed-0568-13-03
10. Soelberg KK, Bonnema SJ, Brix TH, Hegedus L. Risk of malignancy in thyroid incidentalomas detected by 18F-fluorodeoxyglucose positron emission tomography: a systematic review. Thyroid. 2012;22:918-25. https://doi.org/10.1089/thy.2012.0005
11. Pattison DA, Bozin M, Gorelik A, Hofman MS, Hicks RJ, Skandarajah A. 18F-FDG-avid thyroid incidentalomas: the importance of contextual interpretation. J Nucl Med. 2018;59:749-55. https://doi.org/10.2967/jnumed.117.198085
12. Roddy S, Biggans T, Raofi AK, Kanodia A, Sudarshan T, Guntur Ramkumar P. Prevalence of incidental thyroid malignancy on routine 18F-fluorodeoxyglucose PET-CT in a large teaching hospital. Eur J Hybrid Imaging. 2020;4:21. https://doi.org/10.1186/s41824-020-00089-5
13. Buczyńska A, Kościuszko M, Krętowski AJ, Popławska-Kita A. Exploring the clinical utility of DPP-IV and SGLT2 inhibitors in papillary thyroid cancer: a literature review. Front Pharmacol. 2024;15:1323083. https://doi.org/10.3389/fphar.2024.1323083
14. Giovanella L, Milan L, Piccardo A, Bottoni G, Cuzzocrea M, Paone G, et al. Radiomics analysis improves 18FDG PET/CT-based risk stratification of cytologically indeterminate thyroid nodules. Endocrine. 2022;75:202-10. https://doi.org/10.1007/s12020-021-02856-1
15. Zhang X, Ogihara T, Zhu M, Gantumur D, Li Y, Mizoi K, et al. Effect of metformin on 18F-fluorodeoxyglucose uptake and positron emission tomographic imaging. Br J Radiol. 2022;95:20200810. https://doi.org/10.1259/bjr.20200810
16. Sperry BW, Almohtasib Y, Ghaly R, Abdel Jawad M, Sauer AJ, Bateman TM. SGLT2 inhibitor treatment is associated with reduced cardiac glucose metabolism: a matched FDG-PET cohort study. JACC Adv. 2025;4:102016. https://doi.org/10.1016/j.jacadv.2025.102016
17. 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;42:328-54. https://doi.org/10.1007/s00259-014-2961-x.
18. Durante C, Hegedüs L, Czarniecka A, Paschke R, Russ G, Schmitt F, et al. 2023 European Thyroid Association Clinical Practice Guidelines for thyroid nodule management. Eur Thyroid J. 2023;12:e230067. https://doi.org/10.1530/ETJ-23-0067
19. Ulaş Babacan Ö, Hasbek Z. Comparison of 18F-FDG PET/CT data with type 2 diabetes and obesity in breast cancer. Kırıkkale Üniv Tıp Fak Derg. 2025;27:306-11. https://doi.org/10.24938/kutfd.1696624
20. Zhuang H, Codreanu I. Growing applications of FDG PET-CT imaging in non-oncologic conditions. J Biomed Res. 2015;29:189-202. https://doi.org/10.7555/JBR.29.20140081
21. Keskin Y, Bağcı D, Ulucanlar AH, Uçmak G. Beyond the primary tumor: malignancy risk and evaluation strategies for 18F-FDG PET/CT-detected incidentalomas. Mol Imaging Radionucl Ther. 2026;35(1):28-34. https://doi.org/10.4274/mirt.galenos.2025.75768
22. Kim MK, Kim H, Shin JH, Lee JH, Choi JY, Hahn SY. Thyroid incidentalomas detected on 18F-fluorodeoxyglucose positron emission tomography with computed tomography in cancer patients: imaging clues for further evaluation. Eur Radiol. 2026;36:3820-8. https://doi.org/10.1007/s00330-025-12181-0

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260914

Opportunistic computed tomography-based vertebral attenuation for osteoporosis screening

Irfan Atik[1]*, Aibek Tashimbekov[2], Enes Gul[1], Seda Atik[3], Bulent Yildiz[1]

SUMMARY

OBJECTIVE: The aim of this study was to evaluate opportunistic vertebral trabecular attenuation on routine unenhanced abdominal computed tomography for predicting osteoporosis and its relationship with dual-energy X-ray absorptiometry parameters.

METHODS: This retrospective study inc luded 350 patients (mean age 68.5 years) who underwent both dual-energy X-ray absorptiometry and routine unenhanced abdominal computed tomography for non-oncologic clinical indications within a 3-month interval between 2022 and 2025. Patients were grouped into normal (n=119), osteopenia (n=117), and osteoporosis (n=114) based on L1–L4 T-scores. Two blinded radiologists measured Hounsfield Unit values using regions of interest on high-resolution 0.625-mm sagittal-reconstructed computed tomography images of the L1–L4 vertebral bodies. Interobserver agreement was assessed using intraclass correlation coefficients. Diagnostic performance was evaluated using receiver operating characteristic curve analysis.

RESULTS: Lumbar Hounsfield Unit values decreased progressively from normal to osteoporosis groups (p<0.001). Mean L1–L4 attenuation values for normal, osteopenia, and osteoporosis groups were 148.99±5.54, 120.39±3.84, and 78.53±3.36 Hounsfield Unit, respectively (p<0.001). Interobserver reliability was excellent (intraclass correlation coefficient: 0.90–0.94). L1 Hounsfield Unit measurements showed the highest diagnostic performance for predicting osteoporosis (area under the curve: 0.913; 95%CI 0.870–0.955). An L1 cutoff of ≤95.50 Hounsfield Unit yielded 84.60% sensitivity and 86.40% specificity. Mean L1–L4 Hounsfield Unit also showed high accuracy (area under the curve: 0.881) at a ≤95.00 Hounsfield Unit cutoff (sensitivity: 80.80%, specificity: 77.30%).

CONCLUSION: Opportunistic sagittal lumbar Hounsfield Unit measurements, particularly at L1, provide high diagnostic accuracy and reproducibility for identifying osteoporosis on routine abdominal computed tomography without requiring additional imaging or radiation exposure. This protocol may facilitate early opportunistic screening for osteoporosis and support timely referral for dual-energy X-ray absorptiometry assessment.

KEYWORDS: Biomarkers. Bone density. Computed tomography. Osteoporosis.

INTRODUCTION

Osteoporosis (OP) is a common metabolic bone disease characterized by reduced bone mineral density (BMD) and deterioration of bone microarchitecture, resulting in an increased fracture risk[1]. Osteoporotic fractures are associated with substantial mortality, morbidity, and functional impairment, representing a major public health burden[2]. Despite its high prevalence, OP remains frequently undiagnosed in clinical practice, partly because it is often asymptomatic and requires dedicated diagnostic evaluation[3,4]. Opportunistic assessment of BMD from imaging studies performed for other clinical indications may facilitate the identification of at-risk individuals using existing diagnostic data.

Dual-energy X-ray absorptiometry (DXA) is currently the gold standard for OP diagnosis. However, DXA has several limitations. Degenerative changes and vascular calcifications, particularly in older adults, may lead to overestimation of lumbar spine BMD[5,6]. In addition, DXA requires dedicated examinations, resulting in additional healthcare utilization and potential loss of patients during screening and follow-up. Consequently, opportunistic vertebral density assessment using computed tomography (CT) examinations acquired for other indications has emerged as a practical alternative approach[7,8].

Abdominal CT is frequently performed in routine clinical practice and provides valuable information regarding vertebral trabecular bone structure. Vertebral trabecular attenuation can be rapidly quantified in Hounsfield Units (HUs), providing an indirect measure of BMD. Previous studies have demonstrated significant correlations between lumbar vertebral HU values and DXA-derived T-scores[7,8]. Furthermore, CT-based

[1]Sivas Cumhuriyet University, Faculty of Medicine, Department of Radiology – Sivas, Turkey.
[2]Erciyes University, Faculty of Medicine, Department of Radiology – Kayseri, Turkey.
[3]Sivas Cumhuriyet University, Faculty of Medicine, Department of Physical Medicine and Rehabilitation, Division of Rheumatology – Sivas, Turkey.
*Corresponding author: irfanatik_91@hotmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on June 28, 2026. Accepted on August 08, 2026.
Scientific Editor: José Maria Soares Júnior

measurements show high intra- and interobserver reproducibility and require no additional imaging[9,10]. Sagittal CT reconstructions also enable assessment of vertebral compression fractures[11], while low vertebral trabecular attenuation values have been associated with prevalent and future major osteoporotic fractures[12,13].

This study aimed to evaluate the diagnostic value of opportunistic vertebral trabecular attenuation measurements obtained from abdominal CT examinations for osteoporosis detection and to investigate their relationship with DXA-derived BMD parameters.

METHODS

Study design and patient population

This retrospective study was conducted between 2022 and 2025. The study was approved by the Clinical Research Ethics Committee of Cumhuriyet University (Decision No: 2025-12/26, Date: 11.12.2025) and was performed in accordance with the principles of the Declaration of Helsinki. Due to the retrospective design of the study, written informed consent was not obtained from individual patients. The study protocol and informed consent documentation were reviewed and approved by the ethics committee. All patient data were anonymized before analysis.

Abdominal CT examinations had originally been performed for routine non-oncologic clinical indications, including evaluation of abdominal pain, gastrointestinal disorders, hepatobiliary diseases, urinary tract disorders, and follow-up of benign abdominal conditions. Patients who underwent CT for oncologic staging, cancer surveillance, or major traumatic injuries were not included.

Eligible patients were consecutively identified through the institutional Picture Archiving and Communication System (PACS) and electronic medical record system by screening all patients who underwent both abdominal CT and DXA examinations between 2022 and 2025. All consecutive patients fulfilling the predefined inclusion and exclusion criteria were included, regardless of the clinical indication for DXA examination.

Patients who had both DXA and abdominal CT examinations performed within a maximum interval of 3 months were included in the study. A maximum interval of 3 months was considered acceptable, as previous opportunistic CT studies have accepted intervals of up to 6 months between CT and DXA examinations[8]. Therefore, the interval used in the present study was more stringent. Patients with systemic diseases known to affect bone metabolism (including chronic kidney disease, chronic liver disease, endocrine disorders affecting bone metabolism, inflammatory rheumatic diseases, and metabolic bone diseases other than primary osteoporosis), a history of malignancy, secondary causes of osteoporosis, vertebral fixation hardware, or vertebral compression fractures, scoliosis, severe degenerative changes, or other conditions affecting DXA interpretation were excluded. A history of malignancy and relevant comorbidities was verified using the institutional electronic medical record system. In addition, examinations with significant image-degrading artifacts and cases in which the interval between DXA and CT exceeded 3 months were also excluded.

A total of 451 patients were initially identified through the institutional PACS. After applying the predefined exclusion criteria, 15 patients with vertebral compression fractures, 12 with vertebral fixation hardware, 22 with solid malignancies, 12 with significant image artifacts, and 40 with an interval greater than 3 months between DXA and CT examinations were excluded. The final study population consisted of 350 patients.

Dual-energy X-ray absorptiometry and computed tomography evaluation

Bone mineral density measurements were performed using a Quantitative Digital Radiography (QDR) Hologic 4500W DXA system (Hologic Inc., Bedford, MA, USA), with patients examined in the supine position. Lumbar spine (L1–L4) T-scores were obtained from DXA scans. T-scores were calculated using the manufacturer-provided reference database of the Hologic DXA system. Lumbar spine assessment was performed using the L1–L4 region in all patients. Patients were classified according to the World Health Organization (WHO) criteria based on L1–L4 T-scores derived from DXA. Accordingly, patients with a T-score ≥-1 were defined as normal, those with T-scores between -1 and -2.5 were classified as osteopenia, and those with T-scores ≤-2.5 were classified as osteoporosis.

Computed tomography examinations were performed using a 128-detector CT scanner (GE Revolution EVO, Milwaukee, WI, USA) with a standardized abdominal CT protocol. Acquisition parameters included a tube voltage of 120 kVp, automated tube current modulation (100–250 mAs; target noise index, 14), a pitch of 0.984:1, and a gantry rotation time of 0.5 s. Axial datasets were acquired with 1.25-mm detector collimation and reconstructed at 1.0-mm slice thickness using a standard soft-tissue reconstruction kernel (Standard/Medium Smooth). Sagittal multiplanar reformatted images (0.625-mm thickness) were generated

for trabecular bone attenuation measurements and analyzed using the Sectra PACS system (Sectra AB, Linköping, Sweden). Measurements were performed on the L1, L2, L3, and L4 vertebral bodies using equal-sized circular regions of interest (ROIs) with a fixed area of 100 mm^2. For each vertebral body, HU measurements were obtained from a single mid-sagittal reconstructed image. To ensure accurate assessment of trabecular bone, the ROI was placed within the central trabecular portion of the vertebral body while avoiding cortical bone, end-plate regions, focal sclerotic or lytic lesions, and vertebral venous structures. All measurements were independently performed by two radiologists who were blinded to patient clinical data and to each other's measurements. For statistical analysis, the mean values of measurements obtained by both observers were used (Figure 1).

Statistical analysis

Statistical analyses were performed using Statistical Package for the Social Sciences (SPSS) software (Version 23.0; IBM Corp., Armonk, NY, USA). Continuous variables were expressed as mean±standard deviation and categorical variables as frequencies and percentages. Data normality was assessed using the Kolmogorov-Smirnov test. Comparisons among the normal, osteopenic, and osteoporotic groups were performed using one-way analysis of variance (ANOVA) for continuous variables and the chi-square test for categorical variables. Post-hoc pairwise comparisons were conducted using Tukey's honest significant difference test.

Interobserver agreement for HU measurements was assessed using the intraclass correlation coefficient (ICC). Receiver operating characteristic (ROC) curve analysis was performed to evaluate the diagnostic performance of lumbar vertebral HU measurements for osteoporosis prediction, and optimal cutoff values were determined using the Youden index. A $p<0.05$ was considered statistically significant.

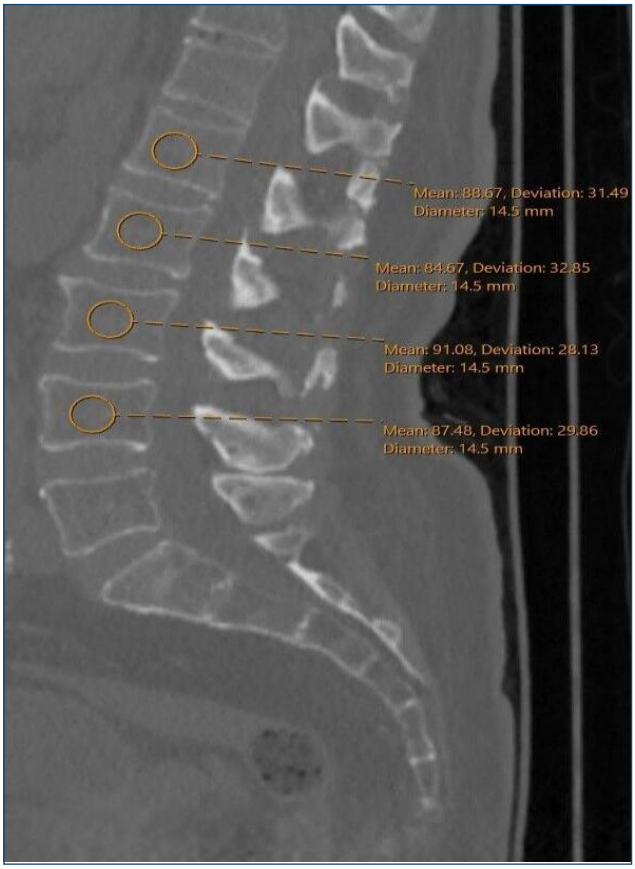


Figure 1. Sagittal reconstructed computed tomography image demonstrating the placement of regions of interest within the trabecular bone of L1–L4 vertebral bodies for Hounsfield Unit measurement.

Table 1. Baseline demographic characteristics and lumbar vertebral computed tomography attenuation values according to dual-energy X-ray absorptiometry -defined bone mineral density groups.

Variable	Normal (n=119)	Osteopenia (n=117)	Osteoporosis (n=114)	p-value
Age (years)	67.8±7.21	68.5±6.92	69.1±7.55	0.42
Female, n (%)	81 (68.1)	80 (68.4)	81 (71.1)	0.89
Male, n (%)	38 (31.9)	37 (31.6)	33 (28.9)	0.91
BMI (kg/m^2)	33.51±2.43	29.03±0.64	28.80±1.03	0.07
L1 HU	154.21±5.85	128.94±3.98	76.56±4.14	**<0.001**
L2 HU	148.50±6.13	122.60±4.28	76.80±3.89	**<0.001**
L3 HU	145.71±6.09	111.93±3.89	78.04±3.77	**<0.001**
L4 HU	147.28±6.34	115.14±4.13	79.95±3.71	**<0.001**
Mean L1–L4 HU	148.99±5.54	120.39±3.84	78.53±3.36	**<0.001**

Continuous variables were compared using one-way analysis of variance (ANOVA), and categorical variables were compared using the chi-square test. Bold values indicate statistically significant results ($p<0.05$). BMI: body mass index; HU: Hounsfield Unit.

Table 2. Diagnostic performance of lumbar vertebral Hounsfield Unit measurements for predicting osteoporosis: receiver operating characteristic curve analysis and optimal cutoff values.

	Cutoff (HU)	Sensitivity (%)	Specificity (%)	AUC	95%CI	p-value
L1 HU	**≤95.50**	84.60	86.40	0.913	0.870–0.955	**<0.001**
L2 HU	**≤92.50**	80.80	72.70	0.861	0.809–0.912	**<0.001**
L3 HU	**≤92.50**	73.10	70.50	0.819	0.759–0.878	**<0.001**
L4 HU	**≤96.50**	72.40	70.50	0.819	0.759–0.878	**<0.001**
Mean L1–L4 HU	**≤95.00**	80.80	77.30	0.881	0.835–0.928	**<0.001**

AUC: area under the receiver operating characteristic curve; CI: confidence interval; HU: Hounsfield Unit; AUC: area under the curve. Bold values indicate statistically significant results (p<0.05) and cutoff HU values.

RESULTS

A total of 350 patients were included in the study; 119 were classified as normal (34%), 117 as osteopenia (33.4%), and 114 as osteoporosis (32.6%). There were no statistically significant differences among the groups in terms of age, sex distribution, or body mass index (BMI) (all p>0.05). Lumbar vertebral HU measurements showed significant differences across all vertebral levels among the groups (p<0.001). Mean L1–L4 HU values demonstrated a progressively decreasing pattern from the normal group to the osteoporotic group, with the highest values observed in the normal group and the lowest in the osteoporosis group (148.99±5.54, 120.39±3.84, and 78.53±3.36, respectively; p<0.001) (Table 1).

Mean L1–L4 HU values showed a significant difference among the groups (p<0.001).

According to Tukey's post-hoc analysis, all pairwise comparisons among groups for L1–L4 vertebral HU values and mean L1–L4 HU values demonstrated statistically significant differences (all p<0.001). The normal group exhibited the highest HU values, whereas the lowest values were observed in the osteoporosis group. The osteopenia group showed intermediate HU values between the normal and osteoporosis groups.

ROC curve analysis demonstrated that lumbar vertebral HU measurements had significant diagnostic performance for predicting osteoporosis (all p<0.001) (Table 2). Among all vertebral levels, the highest diagnostic accuracy was observed for L1 HU measurements (area under the curve [AUC]: 0.913; 95%CI 0.870–0.955). An L1 cutoff value of ≤95.5 HU yielded a sensitivity of 84.6% and a specificity of 86.4% for the diagnosis of osteoporosis. Mean L1–L4 HU values also demonstrated high diagnostic performance (AUC: 0.881; 95%CI 0.835–0.928), with a cutoff value of ≤95 HU providing 80.8% sensitivity and 77.3% specificity. Although L3 and L4 HU measurements showed relatively lower diagnostic performance, all vertebral HU parameters were statistically significant in predicting osteoporosis.

To evaluate the potential confounding effect of BMI, multivariable logistic regression analyses were performed for each vertebral HU parameter (L1–L4 and mean L1–L4). After adjustment for BMI, all vertebral HU measurements remained independently associated with osteoporosis (L1: OR 0.935, 95%CI 0.913–0.958; L2: OR 0.955, 95%CI 0.940–0.971; L3: OR 0.961, 95%CI 0.947–0.975; L4: OR 0.964, 95%CI 0.950–0.977; mean L1–L4: OR 0.941, 95%CI 0.921–0.961; all p<0.001), whereas BMI was not an independent predictor in any of the models (all p>0.05).

Interobserver agreement for lumbar vertebral HU measurements demonstrated excellent reliability between the two radiologists, with ICC values ranging from 0.90 to 0.94.

DISCUSSION

In this study, we evaluated the diagnostic value of opportunistic vertebral trabecular attenuation measurements obtained from unenhanced abdominal CT scans in identifying osteoporosis according to DXA-based WHO criteria. Lumbar vertebral HU values at all levels (L1–L4) and mean L1–L4 HU values showed a significant stepwise decrease across normal, osteopenic, and osteoporotic groups (p<0.001). Furthermore, excellent interobserver agreement (ICC: 0.90–0.94) demonstrates the high interobserver reliability of this method, supporting its use as an objective opportunistic screening tool.

Previous studies have demonstrated that vertebral HU measurements derived from routine abdominal and thoracic CT examinations can be used for opportunistic osteoporosis screening without additional radiation exposure or additional imaging. Pickhardt et al.[8], in a cohort of 1,867 patients, reported a strong correlation between L1 HU values and DXA T-scores, with high sensitivity for osteoporosis detection. Similarly, Pickhardt et al.[7] showed that L1 trabecular attenuation measurements obtained from CT colonography achieved sensitivities approaching 100% for identifying low BMD. These findings

are consistent with the progressive decline in HU values and the high diagnostic performance observed in our study.

ROC analysis demonstrated that the highest diagnostic accuracy for osteoporosis prediction was achieved at the L1 vertebral level (AUC: 0.913). An L1 cutoff value of ≤95.5 HU yielded 84.60% sensitivity and 86.40% specificity. Zhu et al.[14] reported that optimal L1 thresholds generally range from 110 to 135 HU, whereas values ≤100 HU are highly specific for osteoporosis. Therefore, the cutoff identified in our study is consistent with the lower threshold range reported in the literature. Differences among studies may be related to variations in patient demographics and CT acquisition parameters. Lee et al.[15] further demonstrated that every 10-HU decrease in L1 trabecular attenuation is associated with a significantly increased risk of future osteoporotic fractures, highlighting the clinical value of identifying patients with low HU values for earlier DXA evaluation and treatment. The superior diagnostic performance of L1 compared with the mean L1–L4 attenuation may reflect vertebral-level heterogeneity, which could reduce the diagnostic performance of averaged measurements. The optimal cutoff values were determined using the Youden index; however, alternative thresholds may be preferred depending on the intended clinical application.

A major strength of our study is the use of sagittal reconstructed images for HU measurements. Although most opportunistic screening studies have traditionally relied on axial images, recent investigations have compared different imaging planes. Kim et al.[16] reported no significant differences in diagnostic performance among coronal, axial, and sagittal planes, with excellent interobserver agreement for all approaches (ICC>0.90). The authors also emphasized that sagittal and coronal reconstructions may facilitate exclusion of degenerative changes such as osteophytes and end-plate sclerosis while enabling more accurate ROI placement. Consistent with these findings, the sagittal approach used in our study minimized degenerative interference and demonstrated excellent reliability (ICC: 0.90–0.94). Although our findings are generally consistent with previous opportunistic CT studies, the present study adds to the existing literature by demonstrating the diagnostic performance and excellent interobserver reliability of sagittal HU measurements in a Turkish population. These findings support the reproducibility and potential clinical applicability of sagittal plane assessment in routine abdominal CT examinations.

Conventional DXA may overestimate BMD in older individuals because of aortic calcifications, osteophytes, and facet joint degeneration[7,14]. In contrast, CT-based ROI placement within trabecular bone reduces the impact of these artifacts. The marked decline in HU values across the normal, osteopenic, and osteoporotic groups (148.99, 120.39, and 78.53 HU, respectively) indicates that CT attenuation measurements closely reflect DXA-based classifications while providing a more objective assessment.

Although the normal group had a higher mean BMI than the osteopenia and osteoporosis groups, multivariable logistic regression analyses demonstrated that BMI was not an independent predictor of osteoporosis after adjustment for vertebral HU measurements. In contrast, all vertebral HU parameters remained independently associated with osteoporosis. These findings suggest that the diagnostic performance of vertebral HU measurements was not substantially influenced by differences in BMI.

Although the present study demonstrates the diagnostic potential of opportunistic CT-based HU measurements, practical implementation in routine clinical practice will require standardized measurement protocols and, ideally, automated software tools to facilitate workflow integration. In clinical practice, patients with markedly reduced vertebral attenuation values, such as L1 HU values around or below the threshold identified in this study, may be considered for DXA referral based on individual risk factors. Further prospective studies are needed to evaluate implementation feasibility and clinical impact. In addition, the proposed cutoff values should be interpreted in the context of opportunistic screening rather than definitive diagnosis, and positive findings should be confirmed by DXA. Integration of opportunistic bone assessment with other CT-based screening approaches may represent a potential future strategy to maximize the clinical value of routinely acquired CT examinations. Future studies should evaluate automated HU measurement approaches, prospective validation of cutoff values, cost-effectiveness, and integration with electronic health record systems to facilitate broader clinical implementation. From a health economic perspective, opportunistic CT-based HU assessment may reduce healthcare costs by utilizing existing imaging data without additional examinations.

Several limitations should be acknowledged. First, the retrospective single-center design and inclusion of patients who underwent DXA as part of routine clinical care may limit generalizability and introduce selection bias, although consecutive patient inclusion was applied to minimize this effect. Second, intravenous contrast administration may artificially increase vertebral HU values by approximately 10–30 HU[8,16]. To avoid this effect, only non-contrast CT examinations were included; therefore, our findings are directly applicable only to non-contrast CT examinations. Although many abdominal CT studies in routine practice are contrast-enhanced, correction methods

have been proposed for contrast-enhanced CT, but these have not yet been standardized. Future studies incorporating standardized correction factors may improve applicability in these settings. In addition, the proposed HU cutoff values were not internally validated using cross-validation or bootstrapping. Therefore, external validation in independent cohorts is warranted before widespread clinical implementation. Furthermore, although excellent interobserver agreement was demonstrated, intraobserver variability was not assessed. Finally, long-term fracture follow-up data were unavailable.

CONCLUSION

Opportunistic sagittal lumbar vertebral HU measurements obtained from routine abdominal CT scans, particularly at the L1 level and using mean L1–L4 values, provide a reliable, highly reproducible method for opportunistic osteoporosis screening without requiring additional imaging or radiation exposure. Integration of ROI-based HU assessment into routine clinical workflows may facilitate earlier diagnosis and support timely fracture-prevention strategies.

AUTHORS' CONTRIBUTIONS

IA: Conceptualization, Supervision, Validation, Writing – original draft, Writing – review & editing. **AT:** Supervision, Validation, Writing – original draft, Writing – review & editing. **EG:** Validation, Writing – review & editing. **SA:** Supervision, Validation, Writing – review & editing. **BY:** Formal Analysis, 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. Castellani C, Martino E, Scapato P. Osteoporosis: focus on bone remodeling and disease types. BioChem. 2025;5(3):31. https://doi.org/10.3390/biochem5030031
2. Migliorini F, Giorgino R, Hildebrand F, Spiezia F, Peretti GM, Alessandri-Bonetti M, et al. Fragility fractures: risk factors and management in the elderly. Medicina (Kaunas). 2021;57(10):1119. https://doi.org/10.3390/medicina57101119
3. Wang L, Yu W, Yin X, Cui L, Tang S, Jiang N, et al. Prevalence of osteoporosis and fracture in China: the China osteoporosis prevalence study. JAMA Netw Open. 2021;4(8):e2121106. https://doi.org/10.1001/jamanetworkopen.2021.21106
4. Atik İ, Atik S, Gül E, Bulut S. Contribution of lumbar vertebral magnetic resonance ımaging to diagnosis in women with osteoporosis. Turk J Osteoporos. 2022;28(2):77-82. https://doi.org/10.4274/tod.galenos.2021.65037
5. Slart RHJA, Punda M, Ali DS, Bazzocchi A, Bock O, Camacho P, et al. Updated practice guideline for dual-energy X-ray absorptiometry (DXA). Eur J Nucl Med Mol Imaging. 2025;52(2):539-63. https://doi.org/10.1007/s00259-024-06912-6
6. Anupama S, Lim SY, Bolster MB. Updates on the role of DXA in the evaluation and monitoring of osteoporosis. Curr Rheumatol Rep. 2025;27(1):38. https://doi.org/10.1007/s11926-025-01205-9
7. Pickhardt PJ, Lee LJ, Rio AM, Lauder T, Bruce RJ, Summers RM, et al. Simultaneous screening for osteoporosis at CT colonography: bone mineral density assessment using MDCT attenuation techniques compared with the DXA reference standard. J Bone Miner Res. 2011;26(9):2194-203. https://doi.org/10.1002/jbmr.428
8. Pickhardt PJ, Pooler BD, Lauder T, Rio AM, Bruce RJ, Binkley N. Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. Ann Intern Med. 2013;158(8):588-95. https://doi.org/10.7326/0003-4819-158-8-201304160-00003
9. Gerety E-L, Hopper MA, Bearcroft PWP. The reliability of measuring the density of the L1 vertebral body on CT imaging as a predictor of bone mineral density. Clin Radiol. 2017;72(2):177.e9-15. https://doi.org/10.1016/j.crad.2016.09.022
10. Pompe E, Jong PA, Jong WU, Takx RA, Eikendal AL, Willemink MJ, et al. Inter-observer and inter-examination variability of manual vertebral bone attenuation measurements on computed tomography. Eur Radiol. 2016;26(9):3046-53. https://doi.org/10.1007/s00330-015-4145-x
11. Carberry GA, Pooler BD, Binkley N, Lauder TB, Bruce RJ, Pickhardt PJ. Unreported vertebral body compression fractures at abdominal multidetector CT. Radiology. 2013;268(1):120-6. https://doi.org/10.1148/radiol.13121632
12. Graffy PM, Lee SJ, Ziemlewicz TJ, Pickhardt PJ. Prevalence of vertebral compression fractures on routine CT scans according to L1 trabecular attenuation: determining relevant thresholds for opportunistic osteoporosis screening. Am J Roentgenol. 2017;209(3):491-6. https://doi.org/10.2214/AJR.17.17853
13. Lee SJ, Anderson PA, Pickhardt PJ. Predicting future hip fractures on routine abdominal CT using opportunistic osteoporosis screening measures: a matched case-control study. Am J Roentgenol. 2017;209(2):395-402. https://doi.org/10.2214/AJR.17.17820
14. Zhu Y, Triphuridet N, Yip R, Becker BJ, Wang Y, Yankelevitz DF, et al. Opportunistic CT screening of osteoporosis on thoracic and lumbar spine: a meta-analysis. Clin Imaging. 2021;80:382-90. https://doi.org/10.1016/j.clinimag.2021.08.005
15. Lee SJ, Graffy PM, Zea RD, Ziemlewicz TJ, Pickhardt PJ. Future osteoporotic fracture risk related to lumbar vertebral trabecular attenuation measured at routine body CT. J Bone Miner Res. 2018;33(5):860-7. https://doi.org/10.1002/jbmr.3383
16. Kim Y, Kim C, Lee E, Lee JW. Coronal plane in opportunistic screening of osteoporosis using computed tomography: comparison with axial and sagittal planes. Skeletal Radiol. 2024;53(6):1103-9. https://doi.org/10.1007/s00256-023-04525-y

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260920

Digital conscience in midwifery students: a qualitative study

Sumeyye Kose[1], Rojda Bayar Yıldırım[2]*, Sıbel Aydın[3]

SUMMARY

OBJECTIVE: The aim of this study was to explore midwifery students' perceptions of digital technology use in educational and clinical settings, their awareness of ethical issues related to digital environments, and their suggestions for preventing ethical violations.

METHODS: This qualitative exploratory study was conducted in a training and research hospital in Türkiye between March and April 2026. Sixteen third- and fourth-year midwifery students were selected through purposive sampling. Data was collected through face-to-face semi-structured interviews and analyzed using thematic analysis.

RESULTS: Participants reported that digital technologies contributed positively to learning by facilitating access to information, strengthening theoretical knowledge, and improving clinical skills. However, they also emphasized ethical concerns regarding the unauthorized sharing of patient information and photographs, breaches of privacy, and potential legal responsibilities associated with digital technology use. Students highlighted the importance of ethics and privacy education, legal regulations, institutional policies, and individual responsibility in preventing ethical violations in digital environments.

CONCLUSION: Midwifery students generally demonstrated positive attitudes toward digital technology use while recognizing the ethical challenges associated with digital environments. Strengthening digital conscience awareness through education, institutional support, and legal frameworks may promote ethical digital practices, protect patient privacy, and support professional responsibility in midwifery care.

KEYWORDS: Midwifery. Students. Ethics. Privacy. Digital technology.

INTRODUCTION

Digital technologies have become an essential component of healthcare education and clinical practice. Mobile devices, digital learning platforms, artificial intelligence applications, and online resources facilitate access to information, support learning processes, and contribute to the development of clinical competencies among healthcare students[1]. In addition, increasing levels of digital literacy among health sciences students have been associated with improved learning experiences and clinical decision-making skills[2]. Previous studies have demonstrated that innovative digital technologies and e-learning approaches enhance educational outcomes and provide effective support for health professions education[3,4]. Despite these advantages, the integration of digital technologies into educational and clinical environments has introduced important ethical challenges related to privacy, confidentiality, data security, and professional conduct. The use of digital data and online platforms may create concerns regarding personal data sharing, informed consent, and ethical decision-making processes[5]. Furthermore, the unauthorized sharing of patient information, photographs, or clinical experiences through digital platforms may result in privacy violations, professional misconduct, and legal consequences[6,7]. Digital conscience is defined as the ability to use digital technologies in a critical, ethical, and responsible manner. It involves recognizing the risks and opportunities of digital environments, protecting privacy and confidentiality, avoiding misinformation, promoting respectful online communication, and maintaining a balanced use of technology while upholding professional and ethical standards[8-10]. Professional ethics constitutes a fundamental component of midwifery education and practice. Respect for patient dignity, confidentiality, privacy, and professional responsibility are core elements of quality midwifery care. However, previous studies have reported that midwifery students frequently encounter ethical dilemmas during clinical practice and may require additional support in managing ethical challenges[11,12]. Moreover, a recent scoping review highlighted the importance of strengthening digital professionalism among nursing and midwifery students and emphasized the need for

[1]Midwifery Pendik District Health Directorate, Department of Cancer – Istanbul, Türkiye.
[2]Antalya Bilim University, Faculty of Health Sciences, Department of Midwifery – Antalya, Türkiye.
[3]Health Sciences University, Umraniye Training and Research Hospital, Department of Obstetrics and Gynecology – Istanbul, Türkiye.
*Corresponding author: rojdabayar95@outlook.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on April 12, 2026. Accepted on July 19, 2026.
Scientific Editor: José Maria Soares Júnior

greater awareness regarding the ethical use of social media and digital platforms[13]. In this study, digital conscience refers to awareness of ethical responsibility, privacy protection, professional values, and responsible conduct in digital environments. Although digital ethics and professionalism have been studied among healthcare students, evidence regarding digital conscience awareness among midwifery students remains limited. Therefore, this study aimed to explore midwifery students' perceptions of digital technology use, awareness of ethical issues in digital environments, and suggestions for preventing ethical violations.

METHODS

Aim and study design

This study aimed to explore midwifery students' perceptions and experiences regarding the use of digital technologies in education and clinical practice, their awareness of ethical issues associated with digital technology use, and their suggestions for preventing ethical violations. A qualitative exploratory design was adopted to obtain an in-depth understanding of participants' experiences and perspectives.

Research questions

The study was guided by the following research questions:

1. How do midwifery students perceive the use of digital technologies in education and clinical practice?
2. What ethical issues do midwifery students associate with the use of digital technologies?
3. What are midwifery students' recommendations for preventing ethical violations related to digital technology use in clinical settings?

Study variables

As a qualitative study, no dependent or independent variables were defined. The primary areas of inquiry included participants' experiences with digital technologies, perceptions of ethical and privacy-related concerns, awareness of professional responsibility, and suggestions for the prevention of ethical violations.

Study setting and duration

The study was conducted in a training and research hospital in Turkey between March and April 2026.

Population and sample

The study sample was determined using criterion sampling, one of the purposive sampling methods[14]. Participants were third- and fourth-year midwifery students enrolled in the midwifery program of an university in Istanbul, who were undertaking clinical training and met the inclusion criteria. In addition, snowball sampling was used to facilitate participant recruitment, whereby the initial participants referred other eligible students. The sample size was determined according to the principle of data saturation, and data collection continued until no new themes or sub-themes emerged[15]. Accordingly, the study was completed with 16 participants.

Inclusion and exclusion criteria

Eligible participants were third- and fourth-year undergraduate students enrolled in the midwifery program of a university in Istanbul who were actively participating in clinical training, met the inclusion criteria, and provided written informed consent. Students who were not participating in clinical training during the study period, declined to participate, or had communication difficulties were excluded from the study.

Ethical approval

Ethical approval was obtained from the Clinical Research Ethics Committee of Istanbul Umraniye Training and Research Hospital (Decision No: 2026/66, Date: February 12, 2026). The study was conducted in accordance with the principles of the Declaration of Helsinki. Written and verbally informed consent was obtained from all participants prior to data collection.

Data collection instruments

Data was collected using two instruments:

Personal Information Form: Developed by the researchers to collect sociodemographic characteristics and information regarding participants' digital technology use.

Semi-Structured Interview Form: Developed following a review of the relevant literature on digital ethics, digital technology use, and patient privacy[3-5]. The interview guide was reviewed by a panel of 10 experts, including five experts in midwifery and five experts specializing in nursing ethics, to evaluate its content relevance, clarity, and comprehensiveness. Based on their feedback, minor revisions were made to improve the wording and organization of several questions. The final interview guide consisted of open-ended questions exploring perceptions of digital technology use, ethical concerns, professional responsibility, and strategies for preventing ethical violations.

Study procedure

Eligible participants were informed about the purpose and procedures of the study and invited to participate voluntarily. Data was collected through face-to-face semi-structured interviews conducted in a private environment. Interviews lasted

approximately 30–35 min and were audio-recorded with participants' permission. All recordings were transcribed verbatim immediately after the interviews. To protect confidentiality, participants were assigned numerical identifiers (P1–P16).

Statistical analysis

Data was analyzed using the six-phase thematic analysis approach proposed by Braun and Clarke[16,17]. First, audio recordings were transcribed verbatim and reviewed repeatedly to achieve familiarity with the data. Meaningful units were then coded inductively, and similar codes were grouped into themes and sub-themes. To enhance analytical rigor and credibility, two researchers independently reviewed the transcripts and coding structure and discussed emerging themes until consensus was achieved.

The trustworthiness of the findings was supported through expert review of the interview guide, pilot testing with five midwifery students to assess the clarity and comprehensibility of the interview questions, independent coding by two researchers, and the inclusion of direct participant quotations. Reporting of the study was guided by the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist[18].

RESULTS

The study included 16 midwifery students aged 20–23 years. Most participants were in the third or fourth year of study and reported regular use of digital technologies and educational platforms.

Perceptions and experiences of midwifery students regarding the use of digital technologies

Content analysis revealed four main themes and four sub-themes regarding the use of digital technologies in midwifery education and clinical practice. Participants reported that digital technologies reinforced theoretical knowledge, supported clinical skill development, facilitated access to information, and increased professional awareness (Table 1). As P1 stated, "I had never observed a birth before; I first watched it online and prepared myself." Participants also emphasized that simulation applications and digital resources enhanced clinical competence. However, concerns were raised regarding misinformation and the inappropriate use of digital technologies. As P8 noted, "Only experts in the field should share content."

Midwifery students' professional ethical perceptions and experiences regarding the use of digital technologies

Content analysis identified one main theme and three sub-themes related to professional ethical perceptions and experiences regarding digital technologies in midwifery practice. Participants identified privacy violations and unauthorized sharing of patient information and photographs as the primary ethical concerns. They emphasized the importance of protecting confidentiality, obtaining consent, and maintaining professional responsibility in digital environments (Table 2). As P13 stated, "Any sharing related to a patient through digital technologies without the patient's consent constitutes an ethical violation." Participants also highlighted potential legal consequences of privacy breaches, while a few expressed uncertainties regarding the ethical implications of digital technology use.

Solution suggestions for ethical issues in the use of digital technologies in midwifery practice

Analysis revealed four main themes and four sub-themes regarding solution suggestions for ethical issues related to digital

Table 1. Participants' perceptions of digital technologies in midwifery education and professional practice.

Theme	Sub-theme	Participants	Frequency	Sample participant quotation
Contribution to learning	Reinforcement and concretization of theoretical knowledge	1, 3, 4, 8, 10, 12, 13, 14, 15	9	P4: "Visual content makes theoretical knowledge more concrete, and the learning process becomes more effective."
Developing practical skills and clinical competence	Developing practical skills and clinical competence	1, 3, 6, 7, 8, 12, 13, 15	8	P15: "Simulation applications allow us to learn in cases where we cannot observe sufficient cases during clinical practice."
Accessing and dissemination of knowledge	Increasing education and awareness	2, 5, 9, 11, 16	5	P2: "Training can reach a much larger number of people, and people become more aware of these issues."
Limitations in use	Controlled and conscious use of digital technologies	8, 10, 12, 13	4	P8: "Only experts in the field should share content because misinformation can be confusing."

technologies in midwifery practice. Participants emphasized education, legal regulations, institutional policies, and individual responsibility as key strategies for preventing ethical violations. Education was the most frequently recommended approach (Table 3). As P16 stated, "Education should be provided especially in schools on this issue." Participants also highlighted the need for legal sanctions, institutional safeguards, and greater ethical awareness.

DISCUSSION

This study explored midwifery students' perceptions of digital technologies and related ethical issues. Overall, participants viewed digital technologies as valuable tools for learning and professional development while simultaneously recognizing the ethical challenges associated with their use.

Perceptions and experiences regarding the use of digital technologies

Participants reported that digital technologies supported learning by reinforcing theoretical knowledge, facilitating access to information, and enhancing clinical competence. Students emphasized the value of digital resources, simulation applications, and online educational content for clinical preparation. These findings are consistent with studies showing that digital learning tools improve clinical competencies, problem-solving, and decision-making skills among healthcare students[19-21].

Similar findings have shown that digital systems facilitate access to information and support learning during clinical placements[22]. However, some participants expressed concerns regarding misinformation and uncontrolled technology use, highlighting the importance of responsible digital engagement. These findings suggest that digital technologies can enhance midwifery students' preparedness for clinical practice

Table 2. Professional ethical perceptions and experiences regarding the use of digital technologies in midwifery practice.

Theme	Sub-theme	Participants	Frequency	Sample participant quotation
Digital technologies and ethical/privacy issues	Unauthorized sharing of patient information and photographs	1, 3, 5, 7, 8, 10, 11, 12, 13, 14, 15, 16	12	P13: "Any sharing related to a patient through digital technologies without the patient's consent constitutes an ethical violation."
	Professional responsibility and potential legal/ethical risks	4, 6, 8, 15, 16	4	P15: "Violation of patient privacy in the use of digital technologies is an important issue... This may expose us to legal sanctions in our profession."
	Uncertainty and indecision	2, 9	2	P9: "I am undecided about this; I do not fully know whether digital technologies create ethical or professional responsibility issues."

Table 3. Participants' views on solution suggestions for ethical issues in the use of digital technologies.

Theme	Sub-theme	Participants	Frequency	Sample participant quotation
Education and awareness	Increasing ethics and privacy education	1, 2, 4, 6, 7, 8, 9, 10, 13, 16	10	P16: "Education should be provided especially in schools on this issue... In addition, informative training sessions can be organized in hospitals before starting clinical internships."
Legal regulation and supervision	Legal sanctions and monitoring mechanisms	1, 3, 4, 5, 6	5	P4: "Regular inspections can be conducted, and frequent ethics training can be provided."
Institutional regulations and technological measures	Institutional policies and development of digital systems	6, 12, 13, 14, 15	5	P13: "Institutions should introduce specific regulations regarding the use of digital technologies and ensure that this use is carried out within an authorized framework."
Individual responsibility	Empathy, adherence to ethical principles, and conscious behavior	7, 11, 15	3	P11: "I believe that change should start from the individual... we should act in accordance with ethical principles and practice empathy."

when combined with critical thinking and ethical awareness. Therefore, integrating digital literacy and digital professionalism into midwifery education may contribute to the safe, evidence-based, and ethical use of digital technologies in future clinical practice.

Professional ethical perceptions and experiences regarding the use of digital technologies

The most prominent ethical concern was the unauthorized sharing of patient information and photographs. Participants emphasized confidentiality, informed consent, and professional responsibility in digital environments. Similar concerns regarding confidentiality breaches, professional boundaries, and limited awareness of legal responsibilities have been reported previously[23,24]. Therefore, fostering digital conscience during midwifery education may strengthen future midwives' ability to critically evaluate digital information, protect patient confidentiality, and use digital technologies ethically in clinical practice. This interpretation is consistent with the existing literature[25,26].

Solution suggestions for ethical issues in the use of digital technologies

Participants identified education, legal regulation, institutional support, and individual responsibility as key strategies for preventing ethical violations. Education and awareness were the most frequently emphasized recommendations. Previous studies similarly underline the importance of ethics education, digital ethics guidelines, and regulatory frameworks in reducing ethical risks associated with digital technologies[11,27,28]. These findings suggest that the ethical use of digital technologies requires educational, institutional, and regulatory support.

Integrating these strategies into midwifery education through structured curricula, simulation-based learning, institutional digital ethics policies, and legal and ethical training may better prepare students for the responsible use of digital technologies in clinical practice. This approach is consistent with the existing literature highlighting the value of educational technologies and structured educational interventions in strengthening professional competencies among future healthcare professionals[29,30].

Limitations and strengths

The qualitative design, small sample size (n=16), and single-institution setting may limit the transferability of the findings. In addition, because the findings were based on participants' self-reported experiences, responses regarding ethical behavior, patient confidentiality, and professional responsibility may have been influenced by social desirability bias. Nevertheless, the study has several important strengths. The qualitative design enabled an in-depth exploration of midwifery students' perceptions and experiences regarding digital conscience, generating rich and context-specific data that could not be captured through quantitative methods. The use of semi-structured, in-depth interviews facilitated a comprehensive understanding of participants' perspectives, while expert review of the interview guide, pilot testing, and independent coding by two researchers enhanced the credibility and trustworthiness of the findings. Furthermore, reporting the study in accordance with the COREQ guidelines strengthened the methodological transparency and rigor of the research. The findings also contribute to the limited qualitative literature on digital conscience in midwifery education and provide evidence that may inform future educational strategies and professional practice.

CONCLUSION

Midwifery students viewed digital technologies as valuable for learning and clinical skill development but also recognized ethical challenges related to privacy, confidentiality, and professional responsibility. Strengthening digital ethics awareness through education and institutional support may promote the responsible use of digital technologies and help protect patient rights.

ACKNOWLEDGMENTS

The authors thank all subjects who participated in the study.

ETHICAL APPROVAL

Ethical approval for the study was obtained from the Clinical Research Ethics Committee of Umraniye Training and Research Hospital (Decision No: 2026/66, Date: February 12, 2026).

AUTHORS' CONTRIBUTIONS

SK: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Writing – review & editing. **RBY:** Data curation, Formal Analysis, Resources, Visualization, Writing – original draft. **SA:** Methodology, Project administration, , 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. Naciri A, Radid M, Kharbach A, Chemsi G. E-learning in health professions education during the COVID-19 pandemic: a systematic review. J Educ Eval Health Prof. 2021;18:27. https://doi.org/10.3352/jeehp.2021.18.27
2. Aydınlar A, Mavi A, Kütükçü E, Kırımlı EE, Alış D, Akın A, et al. Awareness and level of digital literacy among students receiving health-based education. BMC Med Educ. 2024;24:38. https://doi.org/10.1186/s12909-024-05025-w
3. Reyhan FA, Dağlı E. Investigation of midwifery students' opinions on the digital storytelling method in midwifery education and assessment of their digital stories. Nurse Educ Today. 2024;140:106262. https://doi.org/10.1016/j.nedt.2024.106262
4. Delungahawatta T, Dunne SS, Hyde S, Halpenny L, McGrath D, O'Regan A, et al. Advances in e-learning in undergraduate clinical medicine: a systematic review. BMC Med Educ. 2022;22:711. https://doi.org/10.1186/s12909-022-03773-1
5. Hakimi L, Eynon R, Murphy VA. The ethics of using digital trace data in education: a thematic review of the research landscape. Rev Educ Res. 2021;91:671-717. https://doi.org/10.3102/00346543211020116
6. Zhu X, Hu H, Xiong Z, Zheng T, Li L, Zhang L, et al. Utilization and professionalism toward social media among undergraduate nursing students. Nurs Ethics. 2021;28:297-310. https://doi.org/10.1177/0969733020952105
7. O'Connor S, Odewusi T, Smith PM, Booth RG. Digital professionalism on social media: the opinions of undergraduate nursing students. Nurse Educ Today. 2022;111:105322. https://doi.org/10.1016/j.nedt.2022.105322
8. Universitat Politècnica de Catalunya. Conciencia digital [Internet]. Barcelona: Universitat Politècnica de Catalunya. [cited on 2026 Aug 11]. Available from: https://clickandsafe.upc.edu/es/conciencia-digital
9. Brummer L. Connecting the dots and digits: digital awareness in the context of digital literacy and digital citizenship, and resources for social and digital in-/exclusion. Cogent Soc Sci. 2025;11(1):2572377. https://doi.org/10.1080/23311886.2025.2572377
10. Sánchez Rodríguez E. Ética digital [Internet]. Mexico City: Centro Mexicano para la Filantropía (Cemefi); 2025. [cited on 2026 Jul 11]. Available from: https://www.cemefi.org/wp-content/uploads/2025/01/Ficha-Etica-Digital.pdf
11. Bayrami R, Ghasemzadeh N, Montakhabi Oskuii D. Educational needs of midwifery students in medical ethics: a qualitative study. Nurs Ethics. 2025;32:1129-42. https://doi.org/10.1177/09697330241287836
12. Bygraves D, Wissemann K, Buchanan K. Through the looking glass; midwifery students' experience of ethical issues: a feminist explanatory case study. Nurse Educ Today. 2025;144:106447. https://doi.org/10.1016/j.nedt.2024.106447
13. Gum L, Brown A, Royals R, Matricciani L, Kelly MA. Digital professionalism in preregistration nursing and midwifery students: a scoping review to explore perceptions of professionalism when using social media. Nurse Educ Pract. 2024;80:104128. https://doi.org/10.1016/j.nepr.2024.104128
14. Baltacı A. Nitel araştırmalarda örnekleme yöntemleri ve örnek hacmi sorunsalı üzerine kavramsal bir inceleme. Bitlis Eren Üniversitesi Sosyal Bilimler Dergisi. 2018;7:231-74.
15. Yıldırım A, Şimşek H. Sosyal bilimlerde nitel araştırma yöntemleri. 10th ed. Ankara: Seçkin; 2016.
16. Clarke V, Braun V. Thematic analysis. J Posit Psychol. 2017;12:297-8. https://doi.org/10.1080/17439760.2016.1262613
17. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods. 2017;16:1-13. https://doi.org/10.1177/1609406917733847
18. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research: a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19:349-57. https://doi.org/10.1093/intqhc/mzm042
19. Alsharari AF, Salihu D, Alshammari FF. Effectiveness of virtual clinical learning in nursing education: a systematic review. BMC Nurs. 2025;24:432. https://doi.org/10.1186/s12912-025-03076-y
20. Görücü S, Türk G, Karaçam Z. The effect of simulation-based learning on nursing students' clinical decision-making skills: systematic review and meta-analysis. Nurse Educ Today. 2024;140:106270. https://doi.org/10.1016/j.nedt.2024.106270
21. Clerkin R, Patton D, Moore Z, Nugent L, Avsar P, O'Connor T. What is the impact of video as a teaching method on achieving psychomotor skills in nursing? A systematic review and meta-analysis. Nurse Educ Today. 2022;111:105280. https://doi.org/10.1016/j.nedt.2022.105280
22. Peacock A, Slade C, Brown Wilson C. Nursing and midwifery students' perspectives of using digital systems on placement: a qualitative study. J Adv Nurs. 2022;78:1128-39. https://doi.org/10.1111/jan.15091
23. Oteiza Azcona O, San Juan Moreno A, Villegas Baztán M. Uso de herramientas digitales en la educación prenatal. Ocronos. 2026;9(2):320.
24. Buchanan K, Bygraves D, Ritchie E, Toneycliffe K, Bayes S, Seedhouse D. Woman-centred ethical grid for midwifery education: a content analysis. Nurse Educ Today. 2025;153:106825. https://doi.org/10.1016/j.nedt.2025.106825
25. Kohanová D, Sollárová A, Čakloš M, Zrubcová D, Kolarczyk E. Social media behaviour and patterns of use among nursing students: a systematized review. Nurse Educ Pract. 2025;83:104277. https://doi.org/10.1016/j.nepr.2025.104277
26. Karadaş MM, Koç G, Sayar G, Şahin S, Sevgi T. Student nurses' social media usage, aims of usage and ethical dimension. Gazi Sağlık Bilimleri Dergisi. 2021;6:11-21.
27. Andersson H, Svensson A, Frank C, Rantala A, Holmberg M, Bremer A. Ethics education to support ethical competence learning in healthcare: an integrative systematic review. BMC Med Ethics. 2022;23:29. https://doi.org/10.1186/s12910-022-00766-z
28. Grosman-Rimon L, Wegier P. With advancement in health technology comes great responsibility: ethical and safety considerations for using digital health technology: a narrative review. Medicine. 2024;103:e39136. https://doi.org/10.1097/MD.0000000000039136
29. Silva TJP, Oliveira MFV, Almeida AM, Gonçalves BLS, Dodt RCM, Oriá MOB. Educational technologies on signs of gestational risk and/or labor developed or used to guide pregnant women in the context of Brazilian prenatal care: a scoping review. Rev Esc Enferm USP. 2025;59:e20240433. https://doi.org/10.1590/1980-220X-REEUSP-2024-0433en
30. Fernández Aranda MI. La matrona como evaluadora y gestora de información digital. Matronas Prof. 2016;17(1):20-7.

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260923

Hemorrhoidectomy vs. non-Doppler-guided hemorrhoidal artery ligation

Selçuk Kaya[1]*, Yasin Tosun[1], Nail Can Adıgüzel[1], Önder Altın[1], Mehmet Karahan[1], Cemal Hacıalioğlu[1], Metin Kement[2], Hasan Fehmi Küçük[1]

SUMMARY

OBJECTIVE: Hemorrhoidal disease is one of the most prevalent benign anorectal conditions, and the optimal surgical approach for higher-grade disease remains debated. Conventional excisional hemorrhoidectomy is durable but painful, whereas hemorrhoidal artery ligation with mucopexy is a minimally invasive alternative that aims to reduce morbidity. Although hemorrhoidal artery ligation was originally described as a Doppler-guided procedure, accumulating evidence has questioned whether intraoperative Doppler localization is essential, supporting fixed-position suture ligation as a simpler alternative. We compared conventional hemorrhoidectomy with non-Doppler hemorrhoidal artery ligation combined with mucopexy with respect to perioperative parameters and long-term patient-reported outcomes.

METHODS: This single-center retrospective study included patients who underwent conventional hemorrhoidectomy or non-Doppler hemorrhoidal artery ligation combined with mucopexy for hemorrhoidal disease between June 2023 and December 2024. The hemorrhoidal artery ligation combined with mucopexy technique consisted of a full-thickness Z-suture placed approximately 4 cm proximal to the dentate line for arterial ligation, followed by a running mucopexy of four to six mucosal–submucosal bites at 5-mm intervals down to 5 mm above the dentate line, applied to each of the three classical anatomical hemorrhoidal positions (3, 7, and 11 o'clock). Demographic and perioperative data were obtained by chart review. Postoperative outcomes were assessed by a structured telephone interview conducted blinded to the operative technique, using the Hemorrhoidal Disease Symptom Score, the Short Health Scale for Hemorrhoidal Disease, a 0–10 patient satisfaction score, the Visual Analogue Scale for pain, and the Vaizey incontinence score. Recurrence was defined as the patient's self-reported perception of a symptomatic anal lump and was not confirmed by physical examination. To account for the imbalance in hemorrhoid grade, multivariable logistic regression and a subgroup analysis restricted to grade 3 disease were performed, and effect estimates are reported with 95%CIs.

RESULTS: Of 185 patients operated during the study period, 34 were excluded, leaving 151 patients (62 hemorrhoidectomy, 89 hemorrhoidal artery ligation combined with mucopexy). The groups were comparable in sex, age, American Society of Anesthesiologists class, and smoking status (all $p>0.05$). Hemorrhoid grade distribution differed significantly ($p<0.001$), with grade 2 disease present only in the hemorrhoidectomy group. Hospital stay was shorter after hemorrhoidal artery ligation combined with mucopexy (0.8±0.4 vs. 1.1±0.6 days; $p=0.001$). Over a mean follow-up of approximately 29 months, postoperative bleeding, tenesmus, Visual Analogue Scale pain, Vaizey score, patient satisfaction, and Hemorrhoidal Disease Symptom Score did not differ significantly between groups. Recurrence was numerically higher after hemorrhoidal artery ligation combined with mucopexy (25.8 vs. 12.9%; OR 2.35, 95%CI 0.97–5.68) but did not reach statistical significance ($p=0.083$), and the study was underpowered for this endpoint (post-hoc power 51.8%). The direction of the association persisted after adjustment (aOR 2.21, 95%CI 0.91–5.38) and in the grade 3 subgroup (24.0 vs. 12.8%; OR 2.15, 95%CI 0.73–6.31). Quality of life measured by Short Health Scale for Hemorrhoidal Disease was significantly better (lower) in the hemorrhoidectomy group (6.3±4.3 vs. 8.7±3.9; $p<0.001$).

CONCLUSION: Both techniques achieved comparable symptom control, pain, continence, and satisfaction at long-term follow-up. Non-Doppler hemorrhoidal artery ligation combined with mucopexy offered a shorter hospital stay, while hemorrhoidectomy was associated with better disease-specific quality of life and a numerically lower recurrence rate that persisted after adjustment for hemorrhoid grade but did not reach statistical significance in an underpowered comparison. The results indicate that omission of the Doppler transducer does not compromise the long-term outcomes of hemorrhoidal artery ligation with mucopexy and support an individualized approach to technique selection.

KEYWORDS: Hemorrhoids. Ligation. Recurrence. Quality of life.

[1]Istanbul Kartal Dr. Lutfi Kirdar City Hospital – Istanbul, Türkiye.
[2]VM Medical Park Pendik Hastanesi – Istanbul, Türkiye.
*Corresponding author: selcukkaya_36@hotmail.com

Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on March 25, 2026. Accepted on August 02, 2026.
Scientific Editor: José Maria Soares Júnior

Department of General Surgery, University of Health Sciences, Kartal Dr. Lütfi Kırdar City Hospital, Istanbul, Türkiye.

INTRODUCTION

Hemorrhoidal disease is among the most common benign anorectal disorders encountered in clinical practice, affecting a substantial proportion of the adult population and constituting a frequent reason for surgical referral[1]. The condition arises from the symptomatic enlargement and distal displacement of the normal anal cushions, leading to bleeding, prolapse, pruritus, discomfort, and impaired quality of life[2]. Although conservative and office-based measures are effective for low-grade disease, surgical intervention is generally required for advanced grades and for symptoms refractory to non-operative management[3]. Current guidance reflects this graded approach: the American Society of Colon and Rectal Surgeons clinical practice guidelines recommend excisional hemorrhoidectomy for grade 3 and 4 diseases and for patients who have failed office-based procedures, while recognizing hemorrhoidal artery ligation (HAL) with mucopexy as an acceptable alternative associated with less postoperative pain but a higher likelihood of recurrence[3]. The European Society of Coloproctology guideline similarly positions HAL with mucopexy as an option for grade 2–3 diseases in which prolapse rather than an external component predominates and reserves excisional surgery for advanced and circumferential disease[4]. Neither guideline mandates Doppler guidance as a condition of performing arterial ligation.

Conventional excisional hemorrhoidectomy, performed using open (Milligan-Morgan) or closed (Ferguson) techniques, has long been regarded as the reference standard for advanced hemorrhoidal disease because of its durable results and low recurrence rates[5]. However, excisional surgery is associated with significant postoperative pain, prolonged wound healing, and a measurable risk of complications such as bleeding, urinary retention, and, rarely, anal stenosis or incontinence, all of which may delay return to normal activity[6].

Hemorrhoidal artery ligation (HAL) was introduced as a minimally invasive, non-excisional alternative designed to reduce postoperative morbidity[7]. The technique was originally described using an anoscope incorporating a Doppler transducer to identify and selectively ligate the terminal branches of the superior rectal artery, thereby reducing arterial inflow to the hemorrhoidal plexus[8]. HAL is frequently combined with mucopexy to address the prolapse component, and the resulting reduction in trauma to the somatically innervated anoderm is generally associated with less postoperative pain and faster recovery than excisional surgery[9,10].

Subsequent comparative studies have, however, questioned whether intraoperative Doppler localization is essential to the success of the procedure. The blinded randomized controlled trial by Schuurman et al. found no difference in outcomes between HAL performed with or without the Doppler transducer in grade 2 and 3 hemorrhoidal diseases[11]. Similar conclusions have been reached by Gupta et al. and Avital et al., who reported equivalent outcomes between fixed-position suture ligation and Doppler-guided ligation[12,13]. In parallel, anatomical studies have demonstrated that the terminal branches of the superior rectal artery follow a relatively consistent distribution, supporting fixed-position suture ligation at the typical hemorrhoidal locations as a reasonable surgical strategy. Suture-based hemorrhoidal artery ligation combined with mucopexy at the three classical anatomical positions, performed without Doppler guidance ("non-Doppler HAL with mucopexy," HAL-M), has therefore emerged as a simpler and less equipment-dependent alternative, with non-inferior outcomes demonstrated in randomized comparisons[14,15].

Despite these developments, the comparative effectiveness of non-Doppler HAL-M and conventional excisional hemorrhoidectomy remains incompletely defined, particularly with respect to long-term, validated patient-reported outcome measures (PROMs), which capture the symptomatic and quality-of-life burden of hemorrhoidal disease more meaningfully than clinician-assessed endpoints alone[16].

The aim of the present study was to compare conventional hemorrhoidectomy with non-Doppler HAL combined with mucopexy (HAL-M) with respect to perioperative parameters and long-term patient-reported outcomes, including disease-specific symptom severity, quality of life, pain, continence, recurrence, and overall patient satisfaction, in a real-world surgical cohort.

METHODS

Study design and setting

This study was conducted in the Department of General Surgery, University of Health Sciences, Kartal Dr. Lütfi Kırdar City Hospital, Istanbul, Türkiye. Patients who underwent conventional hemorrhoidectomy or non-Doppler hemorrhoidal artery ligation with mucopexy (HAL-M) for hemorrhoidal disease between June 2023 and December 2024 were included retrospectively. The study was performed in accordance with the principles of the Declaration of Helsinki, and institutional ethics committee approval was obtained. Demographic and perioperative data were retrieved through review of the hospital medical records.

Patient selection

During the study period, a total of 185 patients were operated on using these two techniques in our clinic. Patients with

incomplete perioperative data, those who could not be reached by telephone, and those who declined to participate were excluded. Additional exclusion criteria were age younger than 18 years, prior or subsequent anal or rectal surgery, inflammatory bowel disease, and suspected obstetric anal sphincter injury (OASIS). On the basis of these criteria, 34 patients were excluded, leaving a final study population of 151 patients comprising 62 patients in the conventional hemorrhoidectomy group (Group 1) and 89 patients in the non-Doppler HAL-M group (Group 2).

Surgical technique

In Group 1, conventional excisional hemorrhoidectomy was performed using established open or closed techniques according to surgeon preference and lesion characteristics.

In Group 2, HAL combined with mucopexy was performed without the use of a Doppler probe, using a standardized fixed-position suture technique. With the patient in the lithotomy position, the procedure was applied at each of the three classical anatomical positions of the terminal branches of the superior rectal artery, namely the 3, 7, and 11 o'clock positions, on every anatomical hemorrhoidal packet. At each site, a full-thickness Z-shaped suture was first placed approximately 4 cm proximal to the dentate line using a 40-mm round-bodied 0 polyglactin (Vicryl) suture, with the aim of ligating the hemorrhoidal artery. Continuing distally from this proximal anchor suture, a running mucopexy was then performed by taking 4–6 mucosal–submucosal bites at approximately 5-mm intervals, stopping approximately 5 mm above the dentate line. The mucopexy suture was subsequently tied to the proximal Z-suture, thereby achieving both arterial ligation and plication of the redundant mucosa–submucosa of the hemorrhoidal cushion in a single, integrated suture line. The same procedure was repeated at all three positions, addressing every anatomical hemorrhoidal packet. A schematic representation of the suture technique, showing the proximal Z-shaped arterial ligation suture, the running mucopexy bites, and their relationship to the dentate line, is provided in Figure 1 to aid reproducibility.

Outcome assessment and patient-reported outcome measures

Demographic and perioperative findings, including sex, age, American Society of Anesthesiologists (ASA) physical status

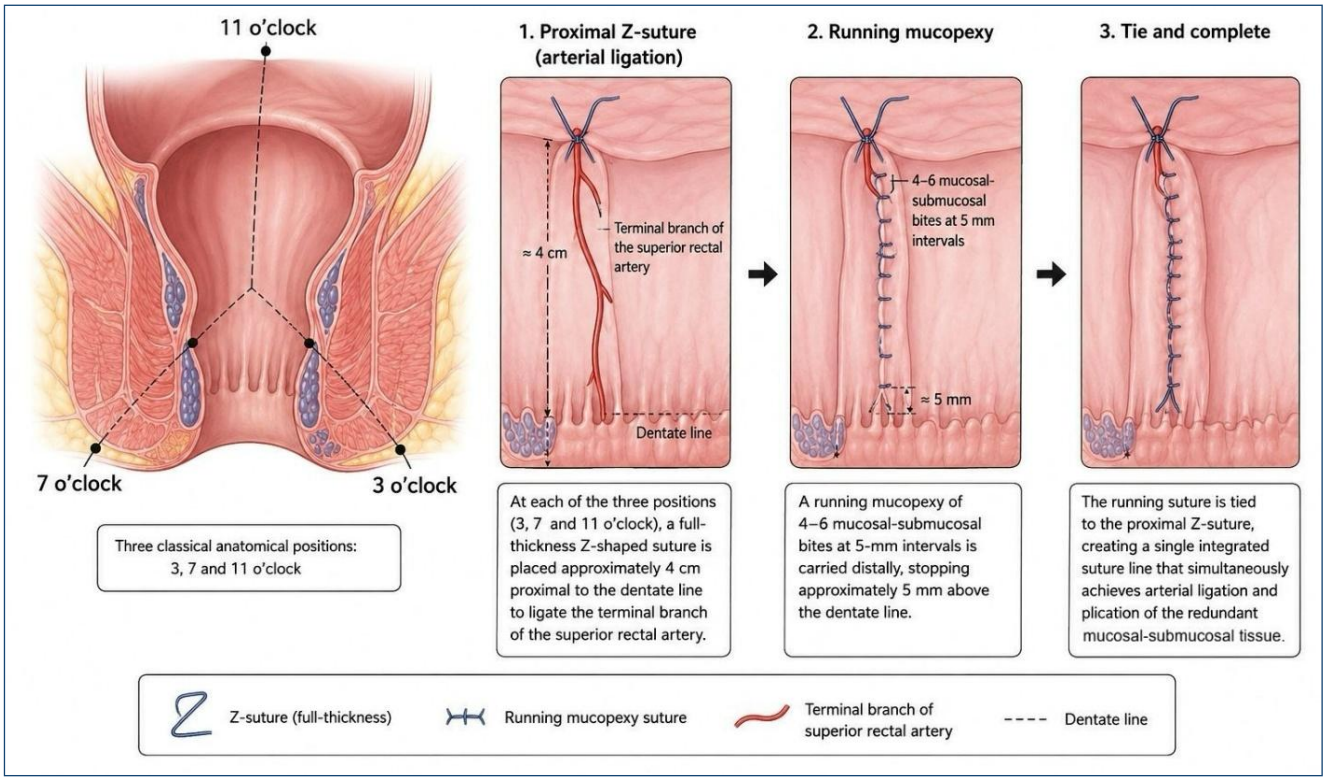


Figure 1. Schematic of the non-Doppler hemorrhoidal artery ligation with mucopexy (HAL-M) technique. At each of the three classical anatomical positions (3, 7, and 11 o'clock), a full-thickness Z-shaped suture is placed approximately 4 cm proximal to the dentate line to ligate the terminal branch of the superior rectal artery. A running mucopexy of 46 mucosal–submucosal bites at 5-mm intervals is then carried distally, stopping approximately 5 mm above the dentate line, and tied to the proximal Z-suture, creating a single integrated suture line that achieves both arterial ligation and plication of the redundant mucosa–submucosa.

classification, smoking status, hemorrhoid grade, and length of hospital stay, were obtained from the medical records. Postoperative findings were determined by structured questionnaires administered to patients by telephone, with the interviewer blinded to the operative technique that had been performed, in order to minimize assessment bias.

Three validated questionnaires were used for reporting PROMs. The severity of hemorrhoidal symptoms was assessed using the Hemorrhoidal Disease Symptom Score (HDSS), a 5-item scale with a possible range from 0 to 20[16]. Health-related quality of life associated with hemorrhoids was evaluated using the Short Health Scale for Hemorrhoidal Disease (SHS-HD), a 4-item scale with a possible range from 0 to 21[16,17]. Validated Turkish-language versions of the HDSS and the SHS-HD were used, and the Vaizey score was administered in its validated Turkish translation; all questionnaires were delivered verbally in Turkish by a single trained interviewer using a standardized script. Patient satisfaction was rated on a numerical scale from 0 to 10. In addition, postoperative pain was assessed using the Visual Analogue Scale (VAS), and anal continence was evaluated using the Vaizey (St. Mark's) incontinence score[18]. Postoperative bleeding, tenesmus, and disease recurrence were also recorded. In this study, recurrence was defined as the patient's self-reported perception of a symptomatic anal lump (pile) causing complaints, as elicited during the telephone interview. This definition was based exclusively on the patient's subjective report and was not confirmed by physical examination. Recurrence was considered present in all patients reporting one or more such episodes during follow-up.

Surgeons and treatment allocation

All operations were performed by four consultant colorectal surgeons, each with more than 10 years of experience in anorectal surgery and each of whom performed both procedures during the study period; no surgeon performed exclusively one technique. All HAL-M procedures followed the standardized fixed-position protocol described above, and the operative steps were identical across surgeons. Because the study was retrospective and non-randomized, allocation to a surgical technique was not protocol-driven. Technique selection was determined primarily by hemorrhoid characteristics and by shared decision-making with the patient: patients with a predominant prolapse component and limited external disease were preferentially offered HAL-M, whereas patients with large external components, circumferential grade 4 disease, associated skin tags or fibrotic piles, and patients with grade 2 disease refractory to office-based treatment were preferentially offered excisional hemorrhoidectomy. Patient preference regarding postoperative pain and expected recovery time also contributed to the decision, as did the availability of day-case operating capacity. Surgeon preference and hospital logistics were not the principal determinants of allocation, but their contribution cannot be excluded and is acknowledged as a source of selection bias.

Statistical analysis

All analyses were performed using SPSS version 26.0 (IBM Corp., Armonk, NY, USA). The normality of continuous variables was assessed using the Shapiro-Wilk test before comparison. Normally distributed continuous variables were expressed as mean±standard deviation and compared using the Student t-test, whereas non-normally distributed variables were compared using the Mann-Whitney U test. Categorical variables were expressed as frequencies and percentages and compared using the Pearson chi-square test or Fisher's exact test, as appropriate. For the primary binary outcome of recurrence, effect estimates are reported as odds ratios (OR), risk ratios (RR), and absolute risk differences with 95%CI in order to convey the precision of the estimates rather than relying on the p-value alone. For continuous outcomes, mean differences with 95%CI and Cohen's d were calculated as measures of effect size. To address the imbalance in hemorrhoid grade between the groups, three additional analyses were prespecified for this revision: (i) a multivariable binary logistic regression model of recurrence including surgical technique, hemorrhoid grade, age, sex, ASA class, and smoking status as covariates, with adjusted OR (aOR) and 95%CI reported; (ii) a subgroup analysis restricted to patients with grade 3 disease, the only grade represented in both groups in sufficient numbers; and (iii) an analysis of covariance for the SHS-HD score adjusted for hemorrhoid grade. A post-hoc power analysis for the observed recurrence proportions was performed using G*Power 3.1 (Heinrich Heine University, Düsseldorf, Germany). A two-sided p-value of less than 0.05 was considered statistically significant.

RESULTS

Baseline and perioperative characteristics

A total of 151 patients were analyzed, including 62 patients (41.1%) in the hemorrhoidectomy group and 89 patients (58.9%) in the non-Doppler HAL-M group. The two groups were comparable with respect to sex distribution (male 74.2 vs. 66.3%; p=0.391), mean age (49.7±14.9 vs. 47.0±13.1 years; p=0.239), ASA classification (p=0.135), and smoking status (27.4 vs. 22.5%; p=0.615). The distribution of hemorrhoid

grade differed significantly between the groups (p<0.001): grade 2 disease was present only in the hemorrhoidectomy group (19.4 vs. 0.0%), grade 3 disease predominated in the HAL-M group (84.3 vs. 62.9%), and grade 4 disease was similarly represented in both groups (17.7 vs. 15.7%). The mean length of hospital stay was significantly shorter after HAL-M (0.8±0.4 vs. 1.1±0.6 days; p=0.001). The baseline and perioperative characteristics of the study groups are summarized in Table 1.

Long-term follow-up and patient-reported outcomes

The mean follow-up duration was comparable between the groups (29.8±5.7 months for hemorrhoidectomy vs. 29.2±4.8 months for HAL-M; p=0.507). At long-term follow-up, there were no statistically significant differences between the groups in postoperative bleeding (16.1 vs. 21.3%; p=0.555), tenesmus (11.3 vs. 7.9%; p=0.668), VAS pain score (0.1±0.4 vs. 0.2±0.4; p=0.371), or Vaizey incontinence score (0.1±0.4 vs. 0.1±0.2; p=0.065). Patient satisfaction was high and similar in both groups (7.6±2.0 vs. 7.5±2.7; p=0.361), as was disease-specific symptom severity measured by HDSS (3.1±3.7 vs. 3.2±2.6; p=0.244).

Recurrence was numerically higher in the HAL-M group than in the hemorrhoidectomy group (23/89, 25.8%, 95%CI 16.7–34.9 vs. 8/62, 12.9%, 95%CI 4.6–21.2), but this difference did not reach statistical significance (p=0.083). The corresponding absolute risk difference was 12.9% (95%CI 0.6–25.3), the unadjusted odds ratio was 2.35 (95%CI 0.97–5.68), and the risk ratio was 2.00 (95%CI 0.96–4.18); the confidence intervals are wide and include values compatible with both no difference and a clinically important excess of recurrence after HAL-M. A post-hoc power calculation based on the observed proportions (25.8 vs. 12.9%, α=0.05, two-sided) yielded a statistical power of 51.8%, and approximately 143 patients per group (286 in total) would have been required to detect a difference of this magnitude with 80% power. The non-significant p-value should therefore be interpreted as reflecting insufficient sample size rather than as evidence of equivalence. In contrast, health-related quality of life measured by the SHS-HD was significantly better in the hemorrhoidectomy group, with lower (more favorable) scores compared with the HAL-M group (6.3±4.3 vs. 8.7±3.9; p<0.001). The postoperative follow-up outcomes are presented in Table 2.

Adjusted and subgroup analyses

Because hemorrhoid grade differed significantly between the groups, recurrence was re-examined in a multivariable binary logistic regression model including surgical technique, hemorrhoid grade, age, sex, ASA class, and smoking status in Table 3. After adjustment, the surgical technique was not an independent predictor of recurrence, although the point estimate remained in the same direction as the unadjusted comparison (HAL-M vs. hemorrhoidectomy: aOR 2.21,

Table 1. Demographic and preoperative characteristics of the study groups.

Variable	Group 1 hemorrhoidectomy (n=62)	Group 2 non-Doppler HAL-M (n=89)	p-value
Gender, n (%)			
Male	46 (74.2)	59 (66.3)	0.391
Female	16 (25.8)	30 (33.7)	
Age, years (mean±SD)	49.7±14.9	47.0±13.1	0.239
ASA classification, n (%)			
ASA 1	11 (17.7)	9 (10.1)	0.135
ASA 2	43 (69.4)	74 (83.1)	
ASA 3	8 (12.9)	6 (6.7)	
Smoking, n (%)	17 (27.4)	20 (22.5)	0.615
Hemorrhoid grade, n (%)			
Grade 2	12 (19.4)	0 (0.0)	**<0.001**
Grade 3	39 (62.9)	75 (84.3)	
Grade 4	11 (17.7)	14 (15.7)	
Hospitalization, days (mean±SD)	1.1±0.6	0.8±0.4	**0.001**

Data are presented as mean±standard deviation or n (%). ASA: American Society of Anesthesiologists physical status; HAL-M: non-Doppler hemorrhoidal artery ligation with mucopexy; SD: standard deviation. p<0.05 are shown in bold.

95%CI 0.91–5.38; p=0.079). Hemorrhoid grade was likewise not an independent predictor (aOR per grade increment 1.66, 95%CI 0.69–3.95; p=0.257), nor were age (aOR 1.01, 95%CI 0.98–1.04; p=0.512), sex (aOR 1.14, 95%CI 0.49–2.66; p=0.762), ASA class (aOR 0.92, 95%CI 0.42–2.02; p=0.836), or smoking status (aOR 1.29, 95%CI 0.53–3.14; p=0.573). The adjusted model therefore did not identify any independent predictor of recurrence, which is consistent with the limited number of recurrence events (n=31) and the correspondingly low events-per-variable ratio; these estimates should be regarded as exploratory.

A subgroup analysis restricted to patients with grade 3 disease, the only grade represented in both groups in adequate numbers (39 hemorrhoidectomy vs. 75 HAL-M), was performed to remove the influence of the grade imbalance. Within this grade-homogeneous subgroup, the direction and magnitude of the difference were preserved but remained non-significant (recurrence 12.8% after hemorrhoidectomy vs. 24.0% after HAL-M; OR 2.15, 95%CI 0.73–6.31; p=0.220), indicating that the recurrence pattern observed in the whole cohort was not solely attributable to differing grade distributions. Propensity score matching was not undertaken because grade 2 disease was entirely absent from the HAL-M group, which produces non-overlapping propensity distributions and precludes meaningful matching on this variable.

The difference in disease-specific quality of life persisted after adjustment for hemorrhoid grade in an analysis of covariance (adjusted mean difference in SHS-HD 2.3 points favoring hemorrhoidectomy, 95%CI 1.0–3.6; p<0.001), compared with an unadjusted mean difference of 2.4 points (95%CI 1.1–3.7; Cohen's d=0.59, a moderate effect). The 0.3-day difference in hospital stay corresponded to a mean difference of 0.30 days (95%CI 0.13–0.47; Cohen's d=0.61).

Figure 2 demonstrates a representative case from the non-Doppler HAL group. Preoperative examination revealed advanced prolapsing hemorrhoidal tissue, whereas postoperative assessment showed successful repositioning of the prolapsed tissue and restoration of the normal anal anatomy.

Table 2. Postoperative follow-up outcomes.

Variable	Group 1 hemorrhoidectomy (n=62)	Group 2 non-Doppler HAL-M (n=89)	p-value
Follow-up, months (mean±SD)	29.8±5.7	29.2±4.8	0.507
Bleeding, n (%)	10 (16.1)	19 (21.3)	0.555
Tenesmus, n (%)	7 (11.3)	7 (7.9)	0.668
VAS pain score (mean±SD)	0.1±0.4	0.2±0.4	0.371
Recurrence, n (%)	8 (12.9)	23 (25.8)	0.083
Vaizey incontinence score (mean±SD)	0.1±0.4	0.1±0.2	0.065
Patient satisfaction (mean±SD)	7.6±2.0	7.5±2.7	0.361
HDSS (mean±SD)	3.1±3.7	3.2±2.6	0.244
SHS-HD (mean±SD)	6.3±4.3	8.7±3.9	**<0.001**

Data are presented as mean±standard deviation or n (%). Follow-up duration was calculated from the date of surgery. HAL-M: non-Doppler hemorrhoidal artery ligation with mucopexy; VAS: Visual Analogue Scale (pain); HDSS: Hemorrhoidal Disease Symptom Score; SHS-HD: Short Health Scale for Hemorrhoidal Disease; Vaizey: Vaizey (St. Mark's) incontinence score. Recurrence was defined as the patient's self-reported perception of a symptomatic anal lump and was not confirmed by physical examination. p<0.05 are shown in bold.

Table 3. Multivariable logistic regression analysis of factors associated with recurrence.

Variable	Adjusted OR	95%CI	p-value
HAL-M vs. hemorrhoidectomy	2.21	0.91–5.38	0.079
Hemorrhoid grade (per increment)	1.66	0.69–3.95	0.257
Age (per year)	1.01	0.98–1.04	0.512
Male sex	1.14	0.49–2.66	0.762
ASA class (per increment)	0.92	0.42–2.02	0.836
Smoking	1.29	0.53–3.14	0.573

OR: odds ratio; CI: confidence interval; HAL-M: non-Doppler hemorrhoidal artery ligation with mucopexy; ASA: American Society of Anesthesiologists physical status. The model included 151 patients with 31 recurrence events; given the low events-per-variable ratio, these estimates should be considered exploratory.



Figure 2. Representative patient treated with the non-Doppler hemorrhoidal artery ligation technique. (A) Preoperative appearance showing prolapsed hemorrhoidal tissue. (B) Immediate postoperative appearance after arterial ligation and mucopexy, demonstrating reduction of prolapse.

DISCUSSION

In this retrospective comparative study of patients undergoing conventional hemorrhoidectomy or non-Doppler hemorrhoidal artery ligation with mucopexy (HAL-M) for symptomatic hemorrhoidal disease, the two techniques achieved broadly comparable long-term results. The groups were well matched for the principal demographic and comorbidity variables, and over a mean follow-up of approximately 29 months there were no significant differences in postoperative bleeding, tenesmus, pain, continence, symptom severity, or patient satisfaction. The main differences favored HAL-M with respect to shorter hospital stays and favored hemorrhoidectomy with respect to disease-specific quality of life and a numerically lower recurrence rate.

The significantly shorter hospital stay observed after HAL-M is consistent with its minimally invasive, non-excisional nature and reflects the reduced tissue trauma and earlier recovery typically reported for this approach[9,10]. The absolute magnitude of this difference is nevertheless small (mean difference 0.30 days, 95%CI 0.13–0.47), and a saving of approximately 7 h is unlikely to be perceived as meaningful by an individual patient. Its relevance is chiefly organizational: with both groups discharged around the first postoperative day, the HAL-M mean of 0.8 days indicates that a majority of these patients were managed as true day cases, whereas the hemorrhoidectomy mean of 1.1 days indicates that most required an overnight stay. At institutional scale, this shifts bed utilization and theater scheduling appreciably, even though the per-patient difference is modest, and it should be interpreted as a health services rather than a patient-level benefit. Notably, despite the difference in technique, postoperative VAS pain scores at long-term follow-up were very low and statistically indistinguishable between the groups, indicating that any early advantage in pain associated with HAL-M converges over time as the excisional wounds heal.

Recurrence remains the principal concern when comparing non-excisional and excisional techniques. In our cohort, recurrence was approximately twice as frequent after HAL-M as after hemorrhoidectomy (25.8 vs. 12.9%), although the difference did not achieve statistical significance, possibly reflecting limited statistical power. This pattern is in keeping with the broader literature, in which excisional hemorrhoidectomy is consistently associated with the lowest recurrence rates, while HAL-based and other minimally invasive procedures may carry a higher risk of recurrent or persistent symptoms, particularly in advanced disease[5,6,10]. The relatively higher proportion of grade 3 disease in the HAL-M group and the presence of grade 2 disease exclusively in the hemorrhoidectomy group reflect real-world patterns of technique selection and represent a potential source of confounding by indication. To address this, we performed a multivariable logistic regression and a subgroup analysis restricted to grade 3 disease. Neither analysis altered the direction of the association: the adjusted odds ratio for recurrence after HAL-M was 2.21 (95%CI 0.91–5.38), and the grade 3 subgroup showed 24.0 versus 12.8% recurrence (OR 2.15, 95%CI 0.73–6.31). This consistency argues against the observed difference being explained by grade imbalance alone, but the CI remain wide and cross unity, so the analyses can neither confirm nor exclude a true difference between the techniques. The post-hoc power of 51.8% confirms that the study was underpowered for this endpoint, and the absence of statistical significance must not be read as evidence of equivalence in recurrence.

An important and somewhat nuanced finding was that disease-specific quality of life, measured with the SHS-HD, was significantly better in the hemorrhoidectomy group, whereas symptom severity measured with the HDSS and overall patient satisfaction did not differ between the groups. This dissociation suggests that, although both procedures effectively control the cardinal symptoms of hemorrhoidal disease, the more definitive removal of pathological tissue achieved by excisional surgery may translate into a measurable advantage in patient-perceived health status[16,17]. The discordance between symptom scores and quality-of-life scores underscores the value of incorporating multiple, complementary PROMs when evaluating outcomes after hemorrhoidal surgery, as no single instrument fully captures the patient experience[16].

A distinctive feature of the present study is that HAL was performed without Doppler guidance, using a standardized suture-ligation and mucopexy technique at the three classical anatomical positions of the superior rectal artery branches. The choice of a non-Doppler approach is supported by accumulating evidence questioning the added value of intraoperative

Doppler localization. The blinded randomized controlled trial by Schuurman et al. demonstrated equivalent clinical outcomes for HAL performed with or without the Doppler transducer in grade 2 and 3 diseases[11], and similar findings have been reported by Gupta et al. and Avital et al.[12,13]. Aigner et al., in a randomized trial of patients with grade 3 disease, found that the addition of Doppler-guided arterial ligation to suture mucopexy reduced early postoperative pain but did not improve recurrence at long-term follow-up[14]. More recently, Perivoliotis et al. demonstrated in a prospective non-inferiority randomized controlled trial that non-Doppler HAL combined with hemorrhoidopexy yielded outcomes comparable to the Doppler-guided technique, further supporting the validity of fixed-position suture-ligation approaches[15]. More recent evidence has reinforced this position. The randomized trial of Alemrajabi et al., published in 2023, compared simple mucopexy with hemorrhoidal arterial ligation performed with and without Doppler guidance and found no advantage for the Doppler-guided approach in short-term outcomes[19]. The most comprehensive assessment to date is the 2025 systematic review and meta-analysis of Tutino et al., which pooled seven randomized controlled trials published between 2008 and 2023 comprising 514 patients and failed to demonstrate any superiority of Doppler-guided over blind ligation with respect to recurrence or postoperative complications; in the subgroup of patients treated with a combined recto-anal repair, which corresponds most closely to the HAL-M technique used here, recurrence and postoperative pain were equivalent between the two approaches[20]. Convergent conclusions were reached by Liew et al. in a 2024 systematic review and meta-analysis of blind versus Doppler-guided HAL-RAR, which likewise supported the non-inferiority of the non-Doppler procedure and emphasized the resulting cost savings for institutions unable to acquire Doppler equipment[21]. Our results, which show satisfactory long-term symptom control, patient satisfaction, and continence after non-Doppler HAL-M, are consistent with this body of evidence and indicate that omission of the Doppler transducer does not compromise the principal long-term clinical outcomes of the procedure. This is particularly relevant for settings where the cost or availability of a Doppler-equipped anoscope may limit the broader uptake of the HAL approach.

A further consideration that must temper the interpretation of our recurrence data is the way recurrence was ascertained. Recurrence was captured solely as the patient's self-reported perception of a symptomatic anal lump during a telephone interview and was never verified by anoscopic or digital examination. Patients frequently cannot distinguish between residual skin tags, a persistent but unchanged prolapsing pile, a thrombosed external hemorrhoid, and genuinely recurrent hemorrhoidal disease, and any of these may be reported as a recurrence. This ascertainment method is likely to have inflated recurrence rates in both arms, and it may have done so differentially: excisional hemorrhoidectomy removes external tissue and skin tags, whereas HAL-M by design leaves the external component in situ, so patients treated with HAL-M may report a palpable lump in the absence of true recurrent disease. The recurrence figures reported here should therefore be regarded as patient-perceived recurrence rather than clinically confirmed recurrence, and they are not directly comparable with series in which recurrence was established by physical examination. Any conclusion regarding the relative durability of the two techniques must await prospective studies with protocol-mandated clinical follow-up.

The very low Vaizey incontinence scores in both groups are reassuring and indicate that neither technique was associated with clinically meaningful deterioration in continence, which is consistent with appropriate technique and patient selection[18]. Similarly, the comparably high satisfaction scores in both groups suggest that, from the patient's perspective, both approaches represent acceptable treatment options when applied to suitable candidates.

Taken together, these results support an individualized approach to technique selection. Non-Doppler HAL-M may be preferred when minimizing hospital stay and early morbidity is prioritized, particularly in lower-grade prolapsing disease, whereas conventional hemorrhoidectomy may offer advantages in disease-specific quality of life and durability of symptom control, especially in advanced disease. Shared decision-making that incorporates disease grade, patient expectations, and tolerance for the possibility of recurrence is therefore essential[3,5].

An economic dimension deserves explicit mention, since one of the stated rationales for omitting the Doppler transducer is cost. We did not perform a formal cost-effectiveness analysis, and detailed institutional cost data were not retrievable retrospectively; the following is therefore a comparison of consumable and capital resources rather than an economic evaluation. Doppler-guided HAL requires a dedicated Doppler-equipped anoscope, a capital item with an acquisition cost in the range of several thousand euros, together with a disposable or reprocessable probe assembly and periodic maintenance. Non-Doppler HAL-M as performed here requires only a standard operating anoscope and three to six sutures of 0 polyglactin per patient, both of which are already present in any general surgical inventory, and adds no capital expenditure. Conventional hemorrhoidectomy similarly requires no dedicated device but consumes an electrosurgical or energy-device handpiece and,

in our practice, was associated with an additional inpatient night in most patients, together with a longer requirement for postoperative analgesia and wound care. In broad terms, therefore, non-Doppler HAL-M avoids both the capital cost of the Doppler system and part of the inpatient cost of excisional surgery, while conventional hemorrhoidectomy avoids the downstream cost of the higher reintervention rate that may follow non-excisional techniques. Whether these opposing cost components favor one approach overall cannot be determined from our data and requires a prospective study incorporating direct and indirect costs, including time to return to work.

This study has several limitations, which we set out explicitly. First, the design is retrospective and single-center, so the analysis is constrained by the completeness of routinely recorded data, and the findings may not be transferable to centers with different case mixes or practice patterns. Second, there was no randomization; allocation to a technique was driven by hemorrhoid characteristics, patient preference, and operating-list availability, which introduces confounding by indication. Third, this allocation process produced a clear selection bias, most visibly in the grade distribution: grade 2 disease occurred exclusively in the hemorrhoidectomy group and grade 3 disease predominated in the HAL-M group, so the groups were not exchangeable at baseline. Fourth, recurrence was ascertained subjectively by telephone and was never confirmed clinically, so the recurrence rates reported here represent patient-perceived rather than objectively verified recurrence and are liable to differential misclassification between an excisional and a non-excisional technique. Fifth, propensity score matching was not feasible: the complete absence of grade 2 patients from the HAL-M group creates non-overlapping propensity distributions, and matching would have discarded a substantial part of the cohort without eliminating the imbalance. Sixth, although we performed multivariable logistic regression, the small number of recurrence events (n=31) yields a low events-per-variable ratio, so the adjusted estimates are exploratory and potentially unstable, and residual confounding by unmeasured factors such as body mass index, bowel habit, straining, and occupational activity cannot be excluded. Seventh, the study was underpowered for the recurrence endpoint (post-hoc power 51.8%), and no formal a priori sample size calculation was performed since the cohort was defined by the study period. Postoperative outcomes were obtained by telephone interview, which, although conducted blinded to the operative technique, relies on patient recall and self-report. In particular, recurrence was defined on the basis of the patient's subjective perception of a symptomatic anal lump and was not verified by physical examination; this may have led to misclassification, since patient-perceived recurrence does not necessarily correspond to objectively confirmed recurrent hemorrhoidal disease, and the reported recurrence rates should therefore be interpreted with caution. The sample size, while adequate to detect large differences, may have been insufficient to demonstrate statistical significance for clinically relevant differences such as recurrence. Although the literature suggests that non-Doppler HAL with mucopexy yields outcomes comparable to Doppler-guided techniques[11-15], our study was not designed to directly compare these two variants. Finally, the absence of detailed cost and early postoperative pain trajectories limits a complete comparison of the two techniques. Prospective, randomized studies with larger samples, standardized grade-matched groups, and longer follow-up, ideally incorporating objective verification of recurrence, are warranted to confirm these findings.

CONCLUSION

Conventional hemorrhoidectomy and non-Doppler HAL with mucopexy provided comparable long-term control of hemorrhoidal symptoms, pain, continence, and patient satisfaction. HAL-M was associated with a significantly shorter hospital stay, whereas hemorrhoidectomy was associated with significantly better disease-specific quality of life and a numerically, though not statistically significant, lower recurrence rate. The omission of the Doppler transducer did not appear to compromise the long-term clinical outcomes of the HAL-with-mucopexy procedure. These findings derive from a retrospective, non-randomized cohort with an imbalanced grade distribution and a recurrence endpoint based on patient self-report, and the comparison of recurrence was underpowered; they should therefore be regarded as hypothesis-generating and require confirmation in prospective randomized studies with objectively verified recurrence. The choice between excisional and minimally invasive techniques should be individualized according to hemorrhoid grade, patient priorities, and acceptance of the relative risk of recurrence, ideally within a shared decision-making framework.

ETHICAL APPROVAL

This study was approved by the Lütfi Kırdar City Hospital Ethics Committee (Date: 11/11/2020; approval number: 514/189/5).

AUTHORS' CONTRIBUTIONS

SK: Investigation, Methodology, Writing – review & editing. **YT:** Data curation, Investigation. **NCA:** Conceptualization,

Data curation. **ÖA:** Investigation, Methodology, Writing – review & editing. **MKa:** Writing – review & editing. **CH:** Formal Analysis, Investigation. **MKe:** Writing – review & editing. **HFK:** Supervision, Validation.

DATA AVAILABILITY STATEMENT

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

REFERENCES

1. Sandler RS, Peery AF. Rethinking what we know about hemorrhoids. Clin Gastroenterol Hepatol. 2019;17(1):8-15. https://doi.org/10.1016/j.cgh.2018.03.020
2. Lohsiriwat V. Hemorrhoids: from basic pathophysiology to clinical management. World J Gastroenterol. 2012;18(17):2009-17. https://doi.org/10.3748/wjg.v18.i17.2009
3. Davis BR, Lee-Kong SA, Migaly J, Feingold DL, Steele SR. The American Society of Colon and Rectal Surgeons Clinical Practice guidelines for the management of hemorrhoids. Dis Colon Rectum. 2018;61(3):284-92. https://doi.org/10.1097/DCR.0000000000001030
4. Tol RR, Kleijnen J, Watson AJM, Jongen J, Altomare DF, Qvist N, et al. European Society of ColoProctology: guideline for haemorrhoidal disease. Colorectal Dis. 2020;22(6):650-62. https://doi.org/10.1111/codi.14975
5. Gallo G, Martellucci J, Sturiale A, Clerico G, Milito G, Marino F, et al. Consensus statement of the Italian society of colorectal surgery (SICCR): management and treatment of hemorrhoidal disease. Tech Coloproctol. 2020;24(2):145-64. https://doi.org/10.1007/s10151-020-02149-1
6. Simillis C, Thoukididou SN, Slesser AA, Rasheed S, Tan E, Tekkis PP. Systematic review and network meta-analysis comparing clinical outcomes and effectiveness of surgical treatments for haemorrhoids. Br J Surg. 2015;102(13):1603-18. https://doi.org/10.1002/bjs.9913
7. Morinaga K, Hasuda K, Ikeda T. A novel therapy for internal hemorrhoids: ligation of the hemorrhoidal artery with a newly devised instrument (Moricorn) in conjunction with a Doppler flowmeter. Am J Gastroenterol. 1995;90(4):610-3. PMID: 7717320
8. Ratto C. THD Doppler procedure for hemorrhoids: the surgical technique. Tech Coloproctol. 2014;18(3):291-8. https://doi.org/10.1007/s10151-013-1062-3
9. Giordano P, Overton J, Madeddu F, Zaman S, Gravante G. Transanal hemorrhoidal dearterialization: a systematic review. Dis Colon Rectum. 2009;52(9):1665-71. https://doi.org/10.1007/DCR.0b013e3181af50f4
10. Pucher PH, Sodergren MH, Lord AC, Darzi A, Ziprin P. Clinical outcome following Doppler-guided haemorrhoidal artery ligation: a systematic review. Colorectal Dis. 2013;15(6):e284-94. https://doi.org/10.1111/codi.12205
11. Schuurman JP, Borel Rinkes IH, Go PM. Hemorrhoidal artery ligation procedure with or without Doppler transducer in grade II and III hemorrhoidal disease: a blinded randomized clinical trial. Ann Surg. 2012;255(5):840-5. https://doi.org/10.1097/SLA.0b013e31824e2bb5
12. Gupta PJ, Kalaskar S, Taori S, Heda PS. Doppler-guided hemorrhoidal artery ligation does not offer any advantage over suture ligation of grade 3 symptomatic hemorrhoids. Tech Coloproctol. 2011;15(4):439-44. https://doi.org/10.1007/s10151-011-0780-7
13. Avital S, Inbar R, Karin E, Greenberg R. Is Doppler ultrasonography essential for hemorrhoidal artery ligation? Tech Coloproctol. 2012;16(4):291-4. https://doi.org/10.1007/s10151-012-0844-3
14. Aigner F, Kronberger I, Oberwalder M, Loizides A, Ulmer H, Gruber L, et al. Doppler-guided haemorrhoidal artery ligation with suture mucopexy compared with suture mucopexy alone for the treatment of Grade III haemorrhoids: a prospective randomized controlled trial. Colorectal Dis. 2016;18(7):710-6. https://doi.org/10.1111/codi.13280
15. Perivoliotis K, Spyridakis M, Zintzaras E, Arnaoutoglou E, Pramateftakis MG, Tepetes K. Non-Doppler hemorrhoidal artery ligation and hemorrhoidopexy combined with pudendal nerve block for the treatment of hemorrhoidal disease: a non-inferiority randomized controlled trial. Int J Colorectal Dis. 2021;36(2):353-63. https://doi.org/10.1007/s00384-020-03768-8
16. Rørvik HD, Styr K, Ilum L, McKinstry GL, Dragesund T, Campos AH, et al. Hemorrhoidal disease symptom score and short health ScaleHD: new tools to evaluate symptoms and health-related quality of life in hemorrhoidal disease. Dis Colon Rectum. 2019;62(3):333-42. https://doi.org/10.1097/DCR.0000000000001234
17. Rørvik HD, Campos AH, Styr K, Ilum L, McKinstry GK, Brandstrup B, et al. Minimal open hemorrhoidectomy versus transanal hemorrhoidal dearterialization: the effect on symptoms: an open-label randomized controlled trial. Dis Colon Rectum. 2020;63(5):655-67. https://doi.org/10.1097/DCR.0000000000001588
18. Vaizey CJ, Carapeti E, Cahill JA, Kamm MA. Prospective comparison of faecal incontinence grading systems. Gut. 1999;44(1):77-80. https://doi.org/10.1136/gut.44.1.77
19. Alemrajabi M, Akbari A, Sohrabi S, Rezazadehkermani M, Moradi M, Agah S, et al. Simple mucopexy and hemorrhoidal arterial ligation with and without Doppler guide: a randomized clinical trial for short-term outcome. Ann Coloproctol. 2023;39(4):351-6. https://doi.org/10.3393/ac.2022.00017.0002
20. Tutino R, Picciariello A, Santarelli M, Simone V, Lobascio P, Cocorullo G, et al. Haemorrhoidal artery ligation: is Doppler guidance useful? A systematic review and meta-analysis of randomized controlled trials. Colorectal Dis. 2025;27(7):e70163. https://doi.org/10.1111/codi.70163
21. Liew AN, Wang J, Chen MZ, Tay YK, Kong JCH. Haemorrhoid artery ligation - recto anal repair (HAL-RAR) blind versus Doppler: a systematic review and meta-analysis. ANZ J Surg. 2024;94(11):2053-61. https://doi.org/10.1111/ans.19258

ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20260932

Association between keratoconus and arterial stiffness: a cross-sectional study

Murat Demirci[1]*, Beste Ozben[1], Esma Ecem Ersoy[2], Serap Karaca[2], Nurten Sayar[1], Urfan Mustafayev[1], Fehim Esen[2], Murat Sunbul[3]

SUMMARY

OBJECTIVE: Keratoconus is a progressive corneal ectasia characterized by stromal thinning and collagen disruption. Despite being primarily considered an ocular disease, recent evidence suggests the possibility of systemic connective tissue involvement. The aim of this study was to evaluate arterial stiffness parameters in patients with keratoconus to explore potential systemic vascular alterations.

METHODS: The study included 43 consecutive patients with keratoconus (mean age: 38.4±12.5 years, 21 male) and 50 healthy controls (mean age: 39.8±7.9 years, 29 male). Arterial stiffness was assessed in all participants using the Mobil-O-Graph system.

RESULTS: The patients with keratoconus had significantly higher pulse wave velocity (6.1±1.1 m/s vs. 5.6±0.5 m/s, p=0.014), augmentation index (21 vs. 12%, p=0.005), peripheral systolic blood pressure (123±12 mmHg vs. 115±8 mmHg, p<0.001), peripheral pulse pressure (46±11 mmHg vs. 39±9, p=0.001), central systolic blood pressure (112±12 mmHg vs. 107±8 mmHg, p=0.016), and central pulse pressure (34±9 mmHg vs. 30±8 mmHg, p=0.037) compared to controls. Pulse wave velocity was significantly correlated with age, systolic blood pressure, body mass index, Kmax, thinnest corneal thickness, and posterior corneal elevation. Keratoconus remained independently associated with pulse wave velocity after adjustment for age, sex, systolic blood pressure, and body mass index (B=0.289, 95%CI 0.083–0.494, p=0.006).

CONCLUSION: Patients with keratoconus had higher pulse wave velocity and augmentation index values than controls, suggesting an association between keratoconus and increased arterial stiffness. These findings highlight the need to consider cardiovascular risk assessment in individuals with keratoconus.

KEYWORDS: Arterial stiffness. Cardiovascular risk. Keratoconus. Pulse wave velocity.

INTRODUCTION

Keratoconus (KC) is characterized by progressive thinning of the cornea and a loss of elasticity, resulting in a conical shape, and is associated with defects in collagen structure[1]. Although the exact etiology remains unclear, environmental factors are thought to contribute to the development of KC[2].

In recent years, more studies have suggested a potential association between KC and systemic diseases[3,4]. KC has also been associated with hereditary connective tissue disorders, including Marfan syndrome and Ehlers-Danlos syndrome[5]. In addition, subclinical myocardial involvement has been demonstrated in patients with KC using echocardiography[6]. Significant alterations in corneal endothelial cell density and morphology have also been reported in advanced KC[7]. Collagen and elastin fibers are essential structural components in maintaining the elasticity of the vascular wall. In patients with KC, connective tissue weakness in the corneal structure may similarly be present in the vascular wall, potentially leading to reduced vascular elasticity and increased arterial stiffness. However, there is a lack of systematic research evaluating arterial stiffness parameters in patients with KC.

The aim of this study was to evaluate arterial stiffness parameters in patients with KC in order to investigate the potential systemic vascular effects of the disease.

METHODS

Study design and participants

This cross-sectional observational study included 43 consecutive patients with a diagnosis of KC for more than 1 year and 50 healthy controls. The diagnosis of KC was confirmed by Scheimpflug corneal topography (Sirius; CSO, Florence, Italy)

[1]Marmara University, Pendik Training and Research Hospital, Department of Cardiology – Istanbul, Turkey.
[2]Istanbul Medeniyet University, Goztepe Prof. Dr. Suleyman Yalcin City Hospital, Department of Ophthalmology – Istanbul, Turkey.
[3]Istinye University, Liv Hospital Topkapi, Department of Cardiology – Istanbul, Turkey.
*Corresponding author: drmuratdemirci@gmail.com
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on May 12, 2026. Accepted on August 04, 2026.
Scientific Editor: José Maria Soares Júnior

after a comprehensive ophthalmic examination. The diagnosis of KC was based on parameters such as K1, K2, Kmean, Kmax, and posterior corneal elevation, together with measurements of the thinnest central corneal thickness and corneal densitometry. KC patients were classified into stages based on the Amsler-Krumeich grading system[8]. Exclusion criteria were the presence of coronary artery disease, congestive heart failure, essential hypertension, diabetes mellitus, hyperlipidemia, atrial fibrillation, chronic kidney disease (with a glomerular filtration rate of less than 60 mL/min/1.73 m^2), chronic inflammatory disease, chronic obstructive pulmonary disease, and history of malignancy. The study was conducted between January 1, 2025, and March 30, 2025.

The baseline characteristics, medical history, comorbidities, and cardiac symptoms of all participants were thoroughly assessed. body mass index (BMI), blood pressure (BP), and 12-lead electrocardiography (ECG) were recorded prior to arterial stiffness. Left ventricular ejection fraction (LVEF) was measured using the modified Simpson's method. Blood samples were taken after an overnight fast of at least 8 h, and lipid profiles and other biochemical parameters were analyzed using standard laboratory methods.

Participants were instructed to abstain from consuming alcohol, coffee, or tea for at least 12 h prior to the assessment in order to minimize confounding factors. Measurements were conducted in a quiet, temperature-controlled room (22–24°C) during the early morning, with participants positioned supine to ensure consistency. Arterial stiffness was assessed using the Mobil-O-Graph (Mobil-O-Graph NG, Stolberg, Germany), a noninvasive oscillometric device commonly used in both clinical and research settings. This system can detect brachial artery pulse wave signals even when cuff pressure exceeds systolic BP by 35 mmHg. The technique analyses pulse wave morphology, which is influenced by arterial stiffness, by capturing the primary pulse wave (the early systolic peak) generated by myocardial contraction and the secondary wave (the late systolic peak), reflected at peripheral branching points. Increased arterial stiffness causes an earlier return of the reflected wave, which raises central systolic pressure and alters wave morphology. The Mobil-O-Graph estimated aortic pulse wave velocity (PWV) and calculated the augmentation index (AIx) (normalized to 75 bpm) by analyzing brachial cuff-derived pressure waveforms. As arterial stiffness increases, vessel rigidity accelerates pulse wave propagation, leading to higher PWV values. The parameters of arterial stiffness were automatically calculated in accordance with current recommendations for noninvasive PWV assessment[9]. Each participant underwent a single arterial stiffness measurement to ensure consistency.

Statistical analysis

Data were analyzed using SPSS software (version 26.0). Normality of continuous variables was assessed using the Shapiro-Wilk test. Normally distributed variables were compared using Student's t-test, whereas non-normally distributed variables were compared using the Mann-Whitney U test. Categorical variables were analyzed using the chi-square test and expressed as counts and percentages n (%). Normally distributed continuous data were presented as mean±standard deviation (SD), whereas non-normally distributed data were presented as median (interquartile range). In KC patients, the association between PWV and other continuous clinical variables was assessed using Pearson or Spearman correlation analysis, based on the distribution of the data. Linear regression analysis was performed to identify independent associations with PWV. A $p<0.05$ was considered statistically significant for all tests.

RESULTS

The study included 43 consecutive patients with KC (mean age: 38.4±12.5 years, 21 male) and 50 healthy controls (mean age: 39.8±7.9 years, 29 male). The demographic and clinical characteristics of the patients and controls are shown in Table 1. There were no significant differences in characteristics regarding age, sex, BMI, LVEF, or biochemical parameters between the patients with KC and the controls.

The arterial stiffness parameters of the KC patients and controls are shown in Table 2. KC patients had significantly higher PWV (6.1±1.1 vs. 5.6±0.5 m/s, p=0.014) and AIx (21 vs. 12%, p=0.005) compared to controls, indicating higher arterial stiffness. KC patients also had higher peripheral and central systolic BP, peripheral and central pulse pressure, while heart rate, cardiac output, and cardiac index were similar between the groups.

In patients with KC, PWV was positively correlated with age (r=0.915, p<0.001), peripheral systolic BP (r=0.595, p<0.001), BMI (r=0.458, p=0.002), posterior corneal elevation (r=0.585, p<0.001), and Kmax (r=0.324, p=0.034), and negatively correlated with thinnest corneal thickness (r=-0.428, p=0.004). The results of the correlation analysis between PWV and clinical variables are presented in Table 3.

Multivariable linear regression analysis showed that KC remained independently associated with PWV after adjustment for age, sex, peripheral systolic BP, and BMI (B=0.289, 95%CI 0.083–0.494, p=0.006). Age (B=0.065, 95%CI 0.056–0.074, p<0.001) and peripheral systolic BP (B=0.034, 95%CI 0.025–0.043, p<0.001) were also independently associated with PWV, whereas sex and BMI were not.

Table 1. Comparison of demographic and clinical characteristics between study groups.

Variables	KC (n=43)	Control (n=50)	p-value
Age (years)	38.4±12.5	39.8±7.9	0.520
Male sex, n (%)	21 (48.8)	29 (58)	0.377
BMI (kg/m^2)	27.2±4.8	26.2±4.3	0.299
LVEF (%)	59.8±6.4	58.6±5.9	0.425
Total-c (mg/dL)	187±42	178.0±68	0.673
LDL-c (mg/dL)	106±33	102±32	0.659
HDL-c (mg/dL)	52±13	50±15	0.759
WBC (×109/L)	7.5±2.3	6.8±1.7	0.223
Hemoglobin (g/dL)	14.1±1.5	13.7±1.3	0.318
Ophthalmological measurements			
K1 (D)	46.21 (43.73–48.42)	NA	NA
K2 (D)	49.37 (46.55–52.94)	NA	NA
Kmean (D)	47.68 (45.27–50.74)	NA	NA
Kmax (D)	57.03 (50.31–61.92)	NA	NA
Thinnest corneal thickness (µm)	432 (395–471)	NA	NA
Corneal densitometry (%)	2.35 (1.18–18.82)	NA	NA
Posterior corneal elevation (µm)	77 (56–135)	NA	NA
Keratoconus stage			
Stage I	23 (53.4%)	NA	NA
Stage II	13 (30.2%)		
Stage III	2 (4.6%)		
Stage IV	5 (11.6%)		

BMI: body mass index; HDL-c: high-density lipoprotein cholesterol; KC: keratoconus; K1, K2, Kmean, and Kmax: keratometric parameters; LDL-c: low-density lipoprotein cholesterol; LVEF: left ventricular ejection fraction; NA: not applicable; Total-c: total cholesterol; WBC: white blood cell count.

Table 2. Comparison of arterial stiffness parameters between study groups.

Parameter	KC (n=43)	Control (n=50)	p-value
Peripheral systolic BP (mmHg)	123±12	115±8	**<0.001**
Peripheral diastolic BP (mmHg)	77±11	76±9	0.650
Average peripheral BP (mmHg)	98±10	94±7	0.065
Peripheral pulse pressure (mmHg)	46±11	39±9	**0.001**
Central systolic BP (mmHg)	112±12	107±8	**0.016**
Central diastolic BP (mmHg)	78±11	77±9	0.521
Central pulse pressure (mmHg)	34±9	30±8	**0.037**
Pulse wave velocity (m/s)	6.1±1.1	5.6±0.5	**0.014**
Augmentation index (%)	21 (13–28)	12 (9–21)	**0.005**
Heart rate (/min)	79±13	75±10	0.134
Cardiac output (L/min)	5.0±0.7	4.8±0.8	0.162
Cardiac index (L/min/m^2)	2.6±0.4	2.5±0.4	0.115

BP: blood pressure; KC: keratoconus; PWV: pulse wave velocity. Bold p-values indicate statistical significance (p<0.05).

Table 3. Correlation analysis between pulse wave velocity and clinical variables in patients with keratoconus.

Variable	r	p-value
Age (years)	0.915	**<0.001**
Peripheral systolic BP (mmHg)	0.595	**<0.001**
BMI (kg/m^2)	0.458	**0.002**
WBC (×10^9/L)	0.267	0.241
Total cholesterol (mg/dL)	0.420	0.083
K1 (D)	0.184	0.238
K2 (D)	0.252	0.102
Kmean (D)	0.225	0.148
Kmax (D)	0.324	**0.034**
Thinnest corneal thickness (µm)	-0.428	**0.004**
Corneal densitometry (%)	0.043	0.785
Posterior corneal elevation (µm)	0.585	**<0.001**

BMI: body mass index; BP: blood pressure; K1, K2, Kmean, Kmax: keratometry readings; PWV: pulse wave velocity; WBC: white blood cell count. Bold p-values indicate statistical significance (p<0.05).

DISCUSSION

Our study shows that PWV and AIx values are significantly higher in KC patients compared with controls, indicating the potential presence of arterial stiffness. In addition, PWV was correlated with Kmax, thinnest corneal thickness, and posterior corneal elevation, suggesting a possible link with arterial stiffness and disease severity. KC was still independently associated with PWV when adjusted for age, sex, systolic BP, and BMI, which are well-known parameters that may affect arterial stiffness. These findings support that KC may not only be an ophthalmological disorder but also may have implications for the cardiovascular system.

Arterial stiffness is a well-established, independent predictor of cardiovascular morbidity and mortality[10]. Increased arterial stiffness has also been reported in systemic diseases such as chronic obstructive pulmonary disease and has been associated with greater carotid atherosclerotic burden[11]. In addition, arterial stiffness and central hemodynamic parameters have been investigated in hypertensive patients with ascending aortic aneurysm, with higher AIx values reported in those with aneurysm[12]. The development of arterial stiffness involves structural changes in the arterial wall, including elastin degradation, collagen accumulation, inflammation, oxidative stress, and extracellular matrix remodeling[13,14]. These processes impair arterial compliance and accelerate pulse wave propagation, leading to higher PWV and central systolic pressure. Similar extracellular matrix remodeling mechanisms may also be involved in the pathogenesis of KC. Altered expression of matrix metalloproteinases (MMPs) and their tissue inhibitors in corneal stromal cells may contribute to collagen remodeling and stromal weakening in KC[15]. The presence of shared extracellular matrix remodeling mechanisms suggests a possible systemic connective tissue dysfunction that could affect both the ocular and vascular structures. Our study demonstrated significantly higher PWV values in KC patients, even after excluding comorbidities such as coronary artery disease, diabetes, and hypertension, which could influence arterial stiffness. Furthermore, despite controlling for classical risk factors such as age, systolic BP, and BMI, PWV remained independently associated with KC, further supporting the existence of a potential connective tissue–vascular axis.

In our study, PWV was significantly correlated with KC-specific corneal parameters, including thinnest corneal thickness, posterior corneal elevation, and Kmax. These corneal alterations reflect underlying stromal weakening, primarily due to disrupted collagen cross-linking and elevated MMP activity—key contributors to the pathogenesis of KC.

Although participants with diagnosed hypertension were excluded, patients with KC had higher peripheral and central systolic BP than controls. Because PWV is influenced by BP at the time of measurement, this difference may partly explain the higher PWV observed in the KC group[16,17]. Peripheral systolic BP was therefore included in the multivariable regression model, and KC remained associated with PWV after adjustment for age, sex, BMI, and peripheral systolic BP. However, residual confounding cannot be completely excluded because both BP and arterial stiffness were assessed at a single visit. Thus, our findings indicate an association between KC and higher PWV, but do not establish a direct causal relationship.

To our knowledge, this is the first study to evaluate arterial stiffness in KC patients using validated noninvasive measurements. By demonstrating increased PWV and AIx in KC patients and showing that these parameters are associated with systemic and corneal structural indicators, our findings provide new insights into the potential systemic vascular involvement in this population. The results suggest that KC may not be solely a localized ocular disorder but could also be linked to systemic connective tissue involvement. Given the paucity of data in this area, our study addresses a significant gap in the literature and lays the groundwork for future research into cardiovascular risk assessment in patients with KC.

LIMITATIONS

This study has several limitations. First, the cross-sectional design prevents us from establishing a causal relationship between KC and arterial stiffness. Second, the relatively small sample size may limit the generalizability of the findings. Third, although

we excluded patients with significant cardiovascular risk factors such as diabetes mellitus, hypertension, and hyperlipidemia, subclinical atherosclerosis or undetected metabolic abnormalities may have influenced the results. Finally, the study did not include long-term follow-up to assess the clinical implications of arterial stiffness in patients with KC.

CONCLUSION

Our study is the first to demonstrate that KC patients have higher PWV and AIx, which may indicate the presence of arterial stiffness. These findings support the hypothesis that KC is not limited to the cornea but may be part of a broader systemic connective tissue disorder with vascular involvement. The significant correlation between PWV and corneal parameters further underlines the structural interplay between ocular and vascular connective tissues. Future longitudinal studies are warranted to assess whether arterial stiffness in KC contributes to cardiovascular risk and to explore the potential benefits of early vascular assessment in this population.

ETHICAL APPROVAL

This study was conducted in accordance with the principles of the Declaration of Helsinki. The study protocol was approved by the Marmara University Ethics Committee on December 24, 2024 (approval no. 09.2024.1542). Written informed consent was obtained from all participants before enrollment.

AUTHORS' CONTRIBUTIONS

MD: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing. **BO:** Conceptualization, Methodology, Supervision, Writing – review & editing. **EEE:** Conceptualization, Investigation, Methodology, Writing – review & editing. **SK:** Conceptualization, Investigation, Methodology, Writing – review & editing. **NS:** Conceptualization, Methodology, Supervision, Writing – review & editing. **UM:** Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Writing – review & editing. **FE:** Conceptualization, Methodology, Supervision, Writing – review & editing. **MS:** Conceptualization, Methodology, 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. Vought R, Greenstein SA, Gelles J, Hersh PS. The pathophysiology of keratoconus. Cornea. 2025;44(2):137-43. https://doi.org/10.1097/ICO.0000000000003585
2. Song M, Fang QY, Seth I, Baird PN, Daniell MD, Sahebjada S. Non-genetic risk factors for keratoconus. Clin Exp Optom. 2023;106(4):362-72. https://doi.org/10.1080/08164622.2022.2062222
3. Lee HK, Jung EH, Cho BJ. Epidemiological association between systemic diseases and keratoconus in a Korean population: a 10-year nationwide cohort study. Cornea. 2020;39(3):348-53. https://doi.org/10.1097/ICO.0000000000002206
4. Moon JY, Lee J, Park YH, Park EC, Lee SH. Incidence of keratoconus and its association with systemic comorbid conditions: a nationwide cohort study from South Korea. J Ophthalmol. 2020;2020:3493614. https://doi.org/10.1155/2020/3493614
5. Barcelo-Canton RH, Ting DSJ, Mehta JS. Genetics of keratoconus: a comprehensive review. Genes (Basel). 2025;16(10):1147. https://doi.org/10.3390/genes16101147
6. Demirci M, Ozben B, Karaca S, Sayar N, Esen F, Tigen K. Assessment of cardiac functions in patients with keratoconus using two-dimensional speckle-tracking echocardiography and three-dimensional echocardiography. Echocardiography. 2025;42(6):e70207. https://doi.org/10.1111/echo.70207
7. Elmassry A, Osman A, Sabry M, Elmassry M, Katkat M, Hatata MY, et al. Corneal endothelial cells changes in different stages of keratoconus: a multi-Centre clinical study. BMC Ophthalmol. 2021;21(1):143. https://doi.org/10.1186/s12886-021-01913-7
8. Deshmukh R, Ong ZZ, Rampat R, Alió Del Barrio JL, Barua A, Ang M, et al. Management of keratoconus: an updated review. Front Med (Lausanne). 2023;10:1212314. https://doi.org/10.3389/fmed.2023.1212314
9. Spronck B, Terentes-Printzios D, Avolio AP, Boutouyrie P, Guala A, Jerončić A, et al. 2024 Recommendations for validation of noninvasive arterial pulse wave velocity measurement devices. Hypertension. 2024;81(1):183-92. https://doi.org/10.1161/HYPERTENSIONAHA.123.21618
10. Vasan RS, Pan S, Xanthakis V, Beiser A, Larson MG, Seshadri S,et al. Arterial stiffness and long-term risk of health outcomes: the framingham heart study. Hypertension. 2022;79(5):1045-56. https://doi.org/10.1161/HYPERTENSIONAHA.121.18776
11. Zhang X, Zhang S, Huang Q, Liu Y, Chang J, Liu P. Comparison of arterial stiffness and ultrasound indices in patients with and without chronic obstructive pulmonary disease. Rev Assoc Med Bras (1992). 2022;68(5):605-9. https://doi.org/10.1590/1806-9282.2021203
12. Arpa A, Özbek M, Güzel T, Polat N. Evaluation of aortic stiffness with an oscillometric arteriograph device in patients with hypertension and ascending aortic aneurysm. Rev Assoc Med Bras (1992). 2026;71(12):e20250481. https://doi.org/10.1590/1806-9282.20250481
13. Herzog MJ, Müller P, Lechner K, Stiebler M, Arndt P, Kunz M, et al. Arterial stiffness and vascular aging: mechanisms, prevention,

and therapy. Signal Transduct Target Ther. 2025;10(1):282. https://doi.org/10.1038/s41392-025-02346-0

14. Wang D, Brady T, Santhanam L, Gerecht S. The extracellular matrix mechanics in the vasculature. Nat Cardiovasc Res. 2023;2(8):718-32. https://doi.org/10.1038/s44161-023-00311-0

15. Berger T, Flockerzi E, Berger M, Chai N, Stachon T, Szentmáry N, et al. Expression of matrix metalloproteinases and their inhibitors in corneal stromal fibroblasts and keratocytes from healthy and keratoconus corneas. Graefes Arch Clin Exp Ophthalmol. 2025;263(2):467-75. https://doi.org/10.1007/s00417-024-06601-y

16. McNally RJ, Boguslavskyi A, Malek R, Floyd CN, Cecelja M, Douiri A, et al. Influence of blood pressure reduction on pulse wave velocity in primary hypertension: a meta-analysis and comparison with an acute modulation of transmural pressure. Hypertension. 2024;81(7):1619-27. https://doi.org/10.1161/HYPERTENSIONAHA.123.22436

17. Kim HL. Arterial stiffness and hypertension. Clin Hypertens. 2023;29(1):31. https://doi.org/10.1186/s40885-023-00258-1

https://doi.org/10.1590/1806-9282.20260070-ER

ERRATUM: Association between lower limb strength, anxiety, and depression symptoms in older adults

In the manuscript: Cruz PP, Sousa RALD, Felício LFF, Monteiro-Junior RS, Cassilhas RC. Association between lower limb strength, anxiety, and depression symptoms in older adults. Rev Assoc Med Bras. 2026;72(7):e20260070. https://doi.org/10.1590/1806-9282.20260070

On page 3, figure 1 caption

Where it reads:

Figure 1. Distribution of 30-Second Chair Stand test performance according to the presence of anxiety and depressive symptoms. Violin plots illustrate the distribution, median, and interquartile range of 30-Second Chair Stand test scores for participants with and without clinically significant anxiety (A) and depressive symptoms (B). Lower 30-Second Chair Stand test performance was significantly associated with anxiety (*p<0.05), whereas the difference for depressive symptoms showed only a marginal trend (p=0.13). These unadjusted group differences visually complement the logistic regression findings reported in Table 3.

It should read:

Figure 1. Distribution of 30-Second Chair Stand test performance according to the presence of anxiety and depressive symptoms. Violin plots illustrate the distribution, median, and interquartile range of 30-Second Chair Stand test scores for participants with and without clinically significant anxiety (A) and depressive symptoms (B). Lower 30-Second Chair Stand test performance was significantly associated with anxiety (*p<0.05), whereas the difference for depressive symptoms showed only a marginal trend (p=0.13).

On page 4

Where it reads:

Table 3. Association between lower limb strength (30-Second Chair Stand test) and anxiety and depressive symptoms.

Variable	Adjusted OR	95%CI	p
Anxiety (GAI≥13), n=74			
LLS—Below vs. normal/above	10.90	1.97–60.33	0.006
Age (per year)	1.18	1.03–1.36	0.019
Sex (male=1)	5.50	0.55–54.60	0.145
Depression (GDS-15≥5), n=59			
LLS—Below vs. normal/above	3.96	1.06–14.74	0.040
Age (per year)	1.05	0.96–1.15	0.323
Sex (male=1)	2.22	0.49–9.98	0.300

OR: odds ratio; CI: confidence interval; GAI: Geriatric Anxiety Inventory; LLS: lower limb strength; GDS-15: Geriatric Depression Scale—15 items.

It should read:

Please disregard Table 3, as it entirely duplicates the content of Table 2.

Roseli Nomura
Editor-in-Chief

CC BY