e-ISSN: 2147-2092

GAZI MEDICAL JOURNAL

galenos yayınevi

medicaljournal.gazi.edu.tr

2026
October
Volume 37 • Issue 4

GAZI MEDICAL JOURNAL

medicaljournal.gazi.edu.tr

Editorial Team

Owner

Uğur Ünal, PhD
Gazi University, Türkiye
E-mail: ugurunal@gazi.edu.tr

Editor in Chief

Mehmet Ali Ergün, MD, PhD
Gazi University, Faculty of Medicine, Department of Medical Genetics, Ankara, Türkiye
E-mail: maliergun@gmail.com

Section Editors and Scientific Editorial Board

Abdullah Özer, MD
Gazi University, Faculty of Medicine, Department of Cardiovascular Surgery, Ankara, Türkiye
E-mail: abdullahozer@gazi.edu.tr

Ahmet Özaslan, MD
Gazi University, Faculty of Medicine, Department of Child and Adolescent Psychiatry, Ankara, Türkiye
E-mail: ahmetozaslan@gazi.edu.tr

Akif Muhtar Öztürk, MD
Gazi University, Faculty of Medicine, Department of Orthopedics and Traumatology, Ankara, Türkiye
E-mail: akifmuhtar@gazi.edu.tr

Alpaslan Şenköylü, MD
Gazi University, Faculty of Medicine, Department of Orthopedics and Traumatology, Ankara, Türkiye
E-mail: senkoylu@gazi.edu.tr

Aylin Sepici Dinçel, MD, PhD
Gazi University, Faculty of Medicine, Department of Biochemistry, Ankara, Türkiye
E-mail: asepici@gazi.edu.tr

Ayşe Meltem Sevgili, MD, PhD
Gazi University, Faculty of Medicine, Department of Physiology, Ankara, Türkiye
E-mail: msevgili@gazi.edu.tr

Burak Sezenöz, MD
Gazi University, Faculty of Medicine, Department of Cardiology, Ankara, Türkiye
E-mail: buraksezenoz@gazi.edu.tr

Cengiz Karakaya, PhD
Gazi University, Faculty of Medicine, Department of Biochemistry, Ankara, Türkiye
E-mail: karakayac@gazi.edu.tr

Çimen Karasu, PhD
Gazi University, Faculty of Medicine, Department of Medical Pharmacology, Ankara, Türkiye
E-mail: karasu@gazi.edu.tr

Gürsel Levent Oktar, MD
Gazi University, Faculty of Medicine, Department of Child and Adolescent Psychiatry, Ankara, Türkiye
E-mail: gloktar@gazi.edu.tr

Hakan Tutar, MD
Gazi University, Faculty of Medicine, Department of Ear, Nose, Throat Diseases, Ankara, Türkiye
E-mail: hakantutar@gazi.edu.tr

Hatice Tuba Atalay, MD
Gazi University, Faculty of Medicine, Department of Ophthalmology, Ankara, Türkiye

GAZI MEDICAL JOURNAL

medicaljournal.gazi.edu.tr

Section Editors and Scientific Editorial Board

E-mail: tubaatalay@gazi.edu.tr

Mehmet Akif Öztürk, MD

Gazi University, Faculty of Medicine, Department of Internal Medicine, Division of Rheumatology, Ankara, Türkiye

E-mail: makifozturk@gazi.edu.tr

Metin Onaran, MD

Gazi University, Faculty of Medicine, Department of Urology, Ankara, Türkiye

E-mail: monaran@gazi.edu.tr

Murat Kekilli, MD

Gazi University, Faculty of Medicine, Department of Internal Medicine, Division of Gastroenterology, Ankara, Türkiye

E-mail: muratkekilli@gazi.edu.tr

Mustafa Arslan, MD

Gazi University, Faculty of Medicine, Department of Anaesthesiology and Reanimation, Ankara, Türkiye

E-mail: mustafaarslan@gazi.edu.tr

Mustafa Sancar Ataç, DMD, PhD

Gazi University, Faculty of Dentistry, Department of Oral and Maxillofacial Surgery, Ankara, Türkiye

E-mail: atac@gazi.edu.tr

Osman Yüksel, MD

Gazi University, Faculty of Medicine, Department of General Surgery, Ankara, Türkiye

Ramazan Karabulut, MD

Gazi University, Faculty of Medicine, Department of Pediatric Surgery, Ankara, Türkiye

E-mail: ramazank@gazi.edu.tr

Serdar Kula, MD

Gazi University, Faculty of Medicine, Department of Pediatrics, Division of Pediatric Cardiology, Ankara, Türkiye

E-mail: kula@gazi.edu.tr

Sezai Leventoğlu, MD

Gazi University, Faculty of Medicine, Department of General Surgery, Ankara, Türkiye

E-mail: sleventoglu@gazi.edu.tr

Sinan Sarı, MD

Gazi University, Faculty of Medicine, Department of Pediatrics, Division of Gastroenterology, Hepatology and Nutrition, Ankara, Türkiye

E-mail: sinansari@gazi.edu.tr

Volkan Medeni, MD, PhD

Gazi University, Faculty of Medicine, Department of Public Health, Ankara, Türkiye

E-mail: volkanmedeni@gazi.edu.tr

Ethical Board

Canan Uluoğlu, MD, PhD

Gazi University, Faculty of Medicine, Department of Medical Pharmacology, Ankara, Türkiye

E-mail: culuoglu@gazi.edu.tr

Nesrin Çobanoğlu, MD, PhD

Gazi University, Faculty of Medicine, Department of Medical Ethics and History of Medicine, Ankara, Türkiye

E-mail: nesrinc@gazi.edu.tr

Nur Aksakal, MD, PhD

Gazi University, Faculty of Medicine, Department of Public Health, Ankara, Türkiye

E-mail: naksakal@gazi.edu.tr

Seçil Özkan, MD, PhD

Gazi University, Faculty of Medicine, Department of Public Health, Ankara, Türkiye

E-mail: ozkans@gazi.edu.tr

Statistical Board

Mustafa N. İlhan, MD, PhD

Gazi University, Faculty of Medicine, Department of Public Health, Ankara, Türkiye

E-mail: mnilhan@gazi.edu.tr

Scientific Advisory Board

Bernd Wollnik

University Medical Center Göttingen, Institute of Human Genetics Heinrich-Düker-Weg 12 37073 Göttingen, Germany

E-mail: bernd.wollnik@med.uni-goettingen.de

GAZI MEDICAL JOURNAL

medicaljournal.gazi.edu.tr

Dan A Zlotolow

Temple University School of Medicine, Shriners Hospital for Children, Department of Orthopaedic Surgery, Philadelphia, PA, USA

Henry Cohen

Gastroenterology Clinic, Montevideo Medical School, Montevideo, Uruguay

Jean-Pierre Michel

Honorary Professor of Medicine (Geneva University, Switzerland) Honorary Professor of Medicine at Limoges University (F) and Beijing University Hospital (CN), Switzerland

Masashi Ohe, MD

Department of Internal Medicine, Japan Community Health Care Organization (JCHO) Hokkaido Hospital, Sapporo, Japan

Mohd Firdaus Mohd Hayati

Universiti Malaysia Sabah, Faculty of Medicine and Health Sciences, Department of Surgery, Kota Kinabalu, Sabah, Malaysia

E-mail: m_firdaus@ums.edu.my

Murat Sincan, MD

National Institute of Dental and Craniofacial Research, National Institutes of Health (NIH), Bethesda, Maryland, USA

Reinhard Büttner

University Hospital Cologne – Institute for Pathology, Center for Integrated Oncology, Cologne, Germany

Thomas Liehr

Universitätsklinikum Jena – Institut für Humangenetik, Jena, Germany

Raja Sabapathy

Department of Plastic and Reconstructive Surgery, Ganga Hospital, Coimbatore, India

E-mail: rajahand@gmail.com

Cagatay Barut, MD, PhD

Bahçeşehir University, Faculty of Medicine, Department of Anatomy, İstanbul, Türkiye

E-mail: cagatay.barut@med.bau.edu.tr

Ebru Evren, MD, PhD

Ankara University, Faculty of Medicine, Department of Medical Microbiology, Ankara, Türkiye

E-mail: evren@ankara.edu.tr

Erkan Yurtcu, PhD

Başkent University, Faculty of Medicine, Department of Medical Biology, Ankara, Türkiye

E-mail: erkanyurtcu@gmail.com

Haktan Bağış Erdem, MD

University of Health Sciences Türkiye, Dr. Abdurrahman Yurtaslan Ankara Oncology Training and Research Hospital, Department of Medical Genetics, Ankara, Türkiye

E-mail: haktanbagis.erdem@sbu.edu.tr

Selahatin Özmen, MD

Koç University, Faculty of Medicine, Department of Plastic, Reconstructive and Aesthetic Surgery, Istanbul, Türkiye

E-mail: profdrsozmen@gmail.com

Hasan Hüseyin Kazan, PhD,

University of Health Sciences Türkiye, Gülhane Faculty of Medicine, Department of Medical Biology, Ankara, Türkiye

E-mail: hasanhuseyinkazan@gmail.com

Please refer to the journal's webpage (https://medicaljournal.gazi.edu.tr/) for "About the Journal" and "Submissions".

The editorial and publication process of Gazi Medical Journal are shaped in accordance with the guidelines of the International Committee of Medical Journal Editors (ICMJE), World Association of Medical Editors (WAME), Council of Science Editors (CSE), Committee on Publication Ethics (COPE), European Association of Science Editors (EASE), and National Information Standards Organization (NISO). The journal is in conformity with the Principles of Transparency and Best Practice in Scholarly Publishing.

Gazi Medical Journal is indexed in Emerging Sources Citation Index, Scopus, Directory of Open Access Journals, EuroPub, Islamic World Science Citation Center, ABCD Index.

The online published articles are freely available on the public internet.

Owner: Musa Yıldız on Behalf of Gazi University

Responsible Manager: Mehmet Ali Ergün

Publisher Contact

Address: Molla Gürani Mah. Kaçamak Sk. No: 21/1
34093 İstanbul, Türkiye
Phone: +90 (539) 307 32 03

E-mail: info@galenos.com.tr/yayin@galenos.com.tr
Web: www.galenos.com.tr
Publisher Certificate Number: 14521

Publication Date: October 2026
e-ISSN: 2147-2092
International scientific journal published quarterly.

GAZI MEDICAL JOURNAL

medicaljournal.gazi.edu.tr

CONTENTS

Original Investigations - Özgün Araştırmalar

461 **Differential Olfactory Dysfunction Indicates Inflammation and Network Degeneration in Multiple Sclerosis**

Multipl Sklerozda Koku İşlev Bozukluğu Paternleri: Enflamasyon ve Nöral Ağ Dejenerasyonu ile İlişkisi

Aslı Aksoy Gündoğdu, Dilcan Kotan, Ayşe Öznur Akidil, Gürkan Kayabaşoğlu, Aytuğ Altundağ

467 **Effects of Oral *Channa striata* and Topical *Centella asiatica* on Serum IL-6 and Wound Tissue TNF-α in a Diabetic Incisional Wound Model in Wistar Rats**

Wistar Sıçanlarında Diyabetik İnsizyonel Yara Modelinde Oral *Channa striata* ve *Topikal Centella* asiatica'nın Serum IL-6 ve Yara Dokusu TNF-α Üzerindeki Etkileri

Dimas Irfan Nabih, Eko Setiawan, Joko Wahyu Wibowo

475 **Investigation of the Level of Awareness and Knowledge of Physicians Regarding the Cancer Susceptibility of Germline *BRCA1* and *BRCA2* Mutations**

Germline *BRCA1* ve *BRCA2* Mutasyonlarının Neden Olduğu Kanser Yatkınlığına Dair Hekimlerin Farkındalık ve Bilgi Düzeylerinin İncelenmesi

Sinan Kolukısaoğlu, Berru Atlıhan, Mustafa Sırdaş Yılmaz, Emre Kağan Dizen, Beyza Doğanay, Şule Altıner

482 **Genetic Diagnosis and Medical Genetics in Turkish Court of Cassation Decisions: A Cross-sectional Decision Analysis**

Türk Yargıtay Kararlarında Genetik Tanı ve Tıbbi Genetik: Kesitsel Bir Karar Analizi

Dilsu Dicle Erkan, Elifcan Taşdelen

490 **Retrospective Re-evaluation of Variants of Uncertain Significance Identified by Clinical Exome Sequencing in Breast Cancer Patients**

Meme Kanseri Hastalarında Klinik Ekzom Dizileme ile Belirlenen Klinik Önemi Belirsiz Varyantların Retrospektif Olarak Yeniden Değerlendirilmesi

Sezen Güntekin Ergün, Feride İffet Şahin

496 **Study Habits and Attitudes Toward Artificial Intelligence Tools Among Anesthesia Trainees in Türkiye: A Descriptive Survey**

Türkiye'deki Anestezi Asistanlarında Çalışma Alışkanlıkları ve Yapay Zeka Araçlarına Yönelik Tutumlar: Tanımlayıcı Bir Anket Çalışması

Çağrı Özdemir, Beyza Büyükgebiz Yeşil, Miray Gözde Özdemir

504 **Reliability of an AI-based Conversational Agent in Pediatric Dentofacial Health Counseling**

Pediatrik Dentofasiyal Sağlık Danışmanlığında Yapay Zeka Tabanlı Konuşma Ajanının Güvenilirliği

Ceren Yapar Gümüş, Süleyman Kutalmış Büyük, Hale Akdemir

511 **The Impact of Clinical Leadership Training on Internal Medicine Residents: A Design-based Quasi-experimental Study**

Klinik Liderlik Eğitiminin İç Hastalıkları Asistanları Üzerindeki Etkisi: Tasarım Tabanlı Yarı Deneysel Bir Çalışma

Mehmet Koca, Özlem Coşkun

GAZI MEDICAL JOURNAL

medicaljournal.gazi.edu.tr

CONTENTS

518 Spectrum of *CYP21A2* Variants: A Retrospective Single-center Study from a Tertiary Healthcare Center in Türkiye

Üçüncü Basamak Sağlık Merkezi Genetik Laboratuvarında *CYP21A2* Genine Yönelik Test Yapılan Olguların Retrospektif Olarak Değerlendirilmesi

Ahmet Kablan, Furkan Baştan, Aysu Yeşim Tezcan, Asuman Bölükbaşı, Şenay Savaş Erdeve, Haktan Bağış Erdem

524 Submandibular Gland Involvement in Head and Neck Cancers: A Single-center Retrospective Study of 175 Cases

Baş ve Boyun Kanserlerinde Submandibular Bez Tutulumu: 175 Olguluk Tek Merkezli Retrospektif Çalışma

Muhittin Akalın, Mehmet Ekrem Zorlu, Emirhan Akyol, Betül Öğüt, Utku Aydil, Metin Yılmaz, Mehmet Düzlü

533 Long-term Outcomes of Kidney Transplant Recipients with Delayed Graft Function: A Single-center Experience

Gecikmiş Greft Fonksiyonu Gelişen Böbrek Nakli Alıcılarının Uzun Dönem Sonuçları: Tek Merkez Deneyimi

Selim Keçeoğlu, Mesut Yavaş, Ramazan Kozan, Aydın Dalgıç, Hakan Sözen

538 Prognostic Significance of LODDS in Right-sided Colon Cancer: A Retrospective Analysis

Sağ Taraflı Kolon Kanserinde LODDS'nin Prognostik Önemi: Retrospektif Bir Analiz

Enes Aksu, Aydın Yavuz, Semih Cihan Yıldırım, Hasan Bostancı, Kürşat Dikmen, Hüseyin Göbüt, Ahmet Çağrı Büyükkasap, Sayın Altıner, Osman Yüksel

544 Persistent or Exacerbated Lymphopenia Predicts Mortality in Kidney Transplant Recipients Hospitalized with COVID-19

COVID-19 Geçiren Böbrek Nakli Alıcılarında Persistan veya Kötüleşen Lenfopeni Mortaliteyi Öngörür

Ali Can Kurtipek, Merve Öztürk, Ali Berkcan Bozdoğan, Ezgi Coşkun Yenigün, Emin Gemcioğlu, Osman İnan, Enes Seyda Şahiner, Oğuzhan Zengin, Selma Karaahmetoğlu, İhsan Ateş, Fatih Dede

551 Seasonal Variation in Histologic and Endoscopic Activity of Pediatric Eosinophilic Esophagitis: A Repeated Endoscopy Cohort

Pediatrik Eozinofilik Özofajitte Histolojik ve Endoskopik Aktivitenin Mevsimsel Değişimi: Tekrarlayan Endoskopi Kohortu

Hakan Öztürk, Ayşe Can, Fatma Eren Kurtipek, Ödül Eğritaş Gürkan, Sinan Sarı, Buket Dalgıç

555 Percutaneous Endoluminal Radiofrequency Ablation Combined with Biliary Stenting for Malignant Biliary Obstruction: A Retrospective Single-center Study

Malign Biliyer Obstrüksiyonda Perkütan Endolüminal Radyofrekans Ablasyon ile Birlikte Biliyer Stentleme: Retrospektif Tek Merkezli Çalışma

Fatih Öncü, Koray Akkan, Baran Önal, Erhan Ilgıt

565 Association Between Regular Postpartum Breastfeeding Counseling and Maternal-infant Bonding: Second-month Postpartum Outcomes

Düzenli Postpartum Emzirme Danışmanlığı ile Anne-bebek Bağlanması Arasındaki İlişki: Postpartum İkinci Ay Sonuçları

Zahide Küçük, Aydan Biri

GAZI MEDICAL JOURNAL

medicaljournal.gazi.edu.tr

CONTENTS

Case Reports - Olgu Sunumları

572 **Malignant Malignant Transformation to Mucinous Adenocarcinoma in Chronic Fistula-in-ano**

Kronik Anal Fistülde Müsinöz Adenokarsinomaya Malign Dönüşüm

Farah Madiehah Abdul Hadi, Khairul Hazim Hamdan, Aswini Bala Krishnan, Firdaus Hayati

576 **A Case of Acromegaly Diagnosed After Pituitary Apoplexy Presenting as Polycystic Ovary Syndrome**

Polikistik Over Sendromu gibi Prezente Olan ve Hipofiz Apopleksisi Sonrasında Tanı Konulan Bir Akromegali Olgusu

Seyit Murat Bayram, Özgür Özçelik, Hüseyin Demirci, Revah Çankaya, Erman Çakal

Literature Review with Cases - Olgularla Literatür İncelemesi

583 **Antibiotic Resistance Risk and Ecosystem**

Antibiyotik Rezistans Riski ve Ekosistem

Evren Doruk Engin, Ayşe Başak Engin, Atilla Engin

596 **Artificial Intelligence in Peroral Endoscopic Myotomy: Current Evidence, Limitations, and Future Directions**

Peroral Endoskopik Miyotomide Yapay Zeka: Güncel Kanıtlar, Kısıtlılıklar ve Geleceğe Yönelik Yaklaşımlar

Ahmed Abdallah Salman, Ahmed Abdallah

Index

2026 Author Index / 2026 Yazar İndeks

2026 Subject Index / 2026 Konu İndeks

Original Investigation | Özgün Araştırma

Gazi Med J 2026;37(4):461-466

DOI: http://dx.doi.org/10.12996/gmj.2026.4592

Differential Olfactory Dysfunction Indicates Inflammation and Network Degeneration in Multiple Sclerosis

Multipl Sklerozda Koku İşlev Bozukluğu Paternleri: Enflamasyon ve Nöral Ağ Dejenerasyonu ile İlişkisi

Aslı Aksoy Gündoğdu[1], Dilcan Kotan[2], Ayşe Öznur Akidil[3], Gürkan Kayabaşoğlu[4], Aytuğ Altundağ[5]

[1]Department of Neurology, Tekirdağ Namık Kemal University, Faculty of Medicine, Tekirdağ, Türkiye
[2]Department of Neurology, Sakarya University, Faculty of Medicine, Sakarya, Türkiye
[3]Department of Otorhinolaryngology, University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital, İstanbul, Türkiye
[4]Alle Clinic, İstanbul, Türkiye
[5]Department of Otorhinolaryngology, Biruni University, Faculty of Medicine, İstanbul, Türkiye

ABSTRACT

Objective: Multiple sclerosis (MS) is an inflammatory, demyelinating and neurodegenerative disease of the central nervous system. Although olfactory dysfunction in MS has been investigated for decades, systematic assessment remains limited. This study aimed to evaluate olfactory function in people with MS (PwMS) in relation to clinical parameters and magnetic resonance imaging (MRI) plaque distribution, using the Sniffin' Sticks test.

Methods: Twenty-eight patients with MS, one participant with clinically isolated syndrome (22 women, 75.9%), and 27 healthy controls were included. Participants with primary olfactory disorders were excluded. Clinical disability was assessed using the Expanded Disability Status Scale (EDSS). Plaque locations on brain MRI scans were recorded. The Sniffin' Sticks test was performed, and threshold, discrimination, identification, and total threshold + discrimination + identification (TDI) scores were obtained for each participant.

Results: Threshold, identification, and TDI scores were significantly lower in PwMS than in controls ($p = 0.021$, 0.004, and 0.005, respectively). Discrimination, identification, and TDI scores were significantly correlated with age ($p = 0.001$, 0.048, and 0.004, respectively) and with disease duration ($p = 0.001$, 0.028, and 0.001, respectively). Olfactory performance was unrelated to sex, smoking, clinical classification, EDSS, or MRI plaque distribution (all $p > 0.05$). A progressive disease course did not influence olfactory measures ($p > 0.05$).

ÖZ

Amaç: Multipl skleroz (MS), merkezi sinir sisteminin enflamatuvar, demiyelinizan ve nörodejeneratif bir hastalığıdır. MS'te koku işlev bozukluğu uzun yıllardır araştırılmasına rağmen koku işlevini sistematik ve çok boyutlu biçimde değerlendiren çalışmaların sayısı halen sınırlıdır. Bu çalışmada, MS'li bireylerde koku işlevinin klinik parametreler ve manyetik rezonans görüntülemedeki (MRG) plak dağılımı ile olan ilişkisinin Sniffin' Sticks testi kullanılarak değerlendirilmesi amaçlandı.

Yöntemler: Çalışmaya ardışık 28 MS hastası, bir klinik izole sendrom katılımcısı (22 kadın, %75,9) ve 27 sağlıklı kontrol dahil edildi. Primer koku bozukluğuna yol açabilecek herhangi bir durumu bulunan katılımcılar çalışma dışında bırakıldı. Klinik dizabilite genişletilmiş engellilik durumu ölçeği (EDSS) ile değerlendirildi. Beyin MRG incelemelerinde saptanan demiyelinizan plakların anatomik lokalizasyonları ayrıntılı olarak kaydedildi. Sniffin' Sticks testi uygulanarak her katılımcının koku eşiği, ayırt etme, tanıma ve toplam eşik + ayırt etme + tanımlama (TDI) skorları ayrı ayrı elde edildi.

Bulgular: Eşik, tanıma ve TDI skorları hasta grubunda kontrollere göre anlamlı düzeyde düşüktü (sırasıyla $p = 0,021$, 0,004 ve 0,005). Ayırt etme, tanıma ve TDI skorları yaş (sırasıyla $p = 0,001$, 0,048 ve 0,004) ve hastalık süresiyle (sırasıyla $p = 0,001$, 0,028 ve 0,001) anlamlı düzeyde ilişkiliydi. Koku performansı cinsiyet, sigara kullanımı, klinik sınıflandırma, EDSS veya MRG plak dağılımı ile ilişkili değildi (tüm $p > 0,05$). Progresif hastalık seyri koku ölçümlerini etkilemedi ($p > 0,05$).

Cite this article as: Aksoy Gündoğdu A, Kotan D, Akidil AÖ, Kayabaşoğlu G, Altundağ A. Differential olfactory dysfunction indicates inflammation and network degeneration in multiple sclerosis. Gazi Med J. 2026;37(4):461-466

Address for Correspondence/Yazışma Adresi: Aslı Aksoy Gündoğdu, Department of Neurology, Tekirdağ Namık Kemal University, Faculty of Medicine, Tekirdağ, Türkiye
E-mail / E-posta: aagundogdu@nku.edu.tr
ORCID ID: orcid.org/0000-0002-6898-0469

Received/Geliş Tarihi: 18.11.2025
Accepted/Kabul Tarihi: 05.09.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

Conclusion: Olfactory dysfunction can be observed across MS subtypes and becomes more pronounced with increasing age and longer disease duration. Threshold impairment may reflect inflammatory activity, whereas reductions in identification and TDI scores likely indicate cumulative neurodegeneration. Multidimensional olfactory testing is simple, accessible, and low-cost. Our findings support its use as a non-invasive complementary biomarker for monitoring long-term neurodegenerative processes in MS.

Keywords: Multiple sclerosis, olfactory dysfunction, Sniffin' Sticks test, threshold, discrimination, identification

ÖZ

Sonuç: Koku bozukluğu MS alt tiplerinde görülebilmekte ve artan yaş ve daha uzun hastalık süresiyle belirginleşmektedir. Eşik bozukluğu enflamatuvar aktiviteyi yansıtabilirken tanıma ve TDI skorlarındaki düşüşler muhtemelen kümülatif nörodejenerasyonu göstermektedir. Çok boyutlu koku değerlendirmesi basit, erişilebilir ve düşük maliyetlidir. Bulgularımız, bu yöntemin MS'de uzun dönemli nörodejeneratif süreçlerin izlenmesinde non-invaziv tamamlayıcı bir biyobelirteç olarak kullanımını desteklemektedir.

Anahtar Sözcükler: Multipl skleroz, koku işlev bozukluğu, Sniffin' Sticks testi, eşik, ayrım, tanıma

INTRODUCTION

Olfactory function is mediated by a neural network consisting of both peripheral and central pathways. Impaired olfactory function negatively affects daily functioning and quality of life (1,2). Olfactory dysfunction is observed in several neurological diseases such as Parkinson's disease, Alzheimer's disease, Huntington disease, motor neuron disease, multiple sclerosis (MS) and migraine (3-7).

MS is a chronic, immune-mediated disease characterized by inflammation, demyelination and neuroaxonal degeneration of the central nervous system (8). Although optic nerve, brainstem, and spinal cord symptoms are well recognized, olfactory dysfunction in MS remains frequently underdiagnosed (6,9). Histopathological evidence demonstrates demyelination and axonal degeneration in the olfactory bulb and tract in approximately 70% of MS brains (10).

The prevalence of olfactory dysfunction in MS ranges from 30% to 75% and has been associated with longer disease duration, higher Expanded Disability Status Scale (EDSS) scores and greater plaque load particularly within frontal and temporal regions (6,7,11,12). Recent studies show that olfactory impairment is more pronounced in progressive MS and correlates with clinical worsening and disability progression (6,7,13).

Emerging evidence indicates that olfactory impairment reflects central network-level neurodegeneration, particularly involving orbitofrontal, limbic and associative cortices (14,15). Longitudinal magnetic resonance imaging (MRI) studies have shown that a decline in olfactory performance parallels structural degeneration in these regions and that orbitofrontal cortex atrophy is strongly linked to impaired odor identification (16).

Various olfactory assessment tools have been used in MS, including the University of Pennsylvania smell identification test, the brief smell identification test, the Connecticut Chemosensory clinical research center test, and the Sniffin' Sticks test (12,17-19). The Sniffin' Sticks test provides a three-domain evaluation (threshold, discrimination, identification) and has been validated in the Turkish population (20).

Furthermore, recent clinical work supports the Sniffin' Sticks test as a practical clinical marker in MS monitoring (13,16). This study aimed to evaluate olfactory function in MS using the Sniffin' Sticks test and to determine relationships between olfactory function and disease duration, EDSS score, and MRI plaque distribution.

MATERIALS AND METHODS

Patient Selection

This study included 29 patients aged 18-65 years, comprising 28 patients with MS and one participant with clinically isolated syndrome (CIS), classified according to the 2017 revised McDonald criteria (21) together with 27 healthy controls. Patients were recruited from the MS outpatient clinic of the Department of Neurology, and healthy controls from the Department of Otorhinolaryngology at patients were recruited from the MS outpatient clinic of the Department of Neurology, and healthy controls from the outpatient clinic of the Department of Otorhinolaryngology at Sakarya University Training and Research Hospital. The control group consisted of individuals without neurological disease or known olfactory dysfunction. The study was approved by the Institutional Ethics Committee of Sakarya University, Faculty of Medicine (ethical approval number: E-16214662-050.01.04-28539-97, date: 21.04.2021), and written informed consent was obtained from all participants.

Exclusion criteria included subjective olfactory complaints, nasal obstruction, history of nasal surgery, nasal polyps, rhinitis, paranasal sinusitis, recent upper respiratory tract infection, head trauma, toxic exposures, smoking smoking >20 cigarettes/day cigarettes/day, pregnancy, hormone replacement therapy, diabetes mellitus, hypertension, major neurological or psychiatric disorders, severe cognitive impairment, Alzheimer's disease, and Parkinson's disease. Patients who had experienced an MS relapse or received corticosteroid treatment within 30 days of olfactory testing were also excluded.

Demographic and clinical data, including age, sex, smoking status, relapse history, and medication use, were recorded. Disease-related variables included disease duration and clinical classification: CIS, relapsing-remitting MS (RRMS), primary progressive MS (PPMS), or secondary progressive MS (SPMS). Neurological examinations were performed by the same neurologist. Disability level was assessed using the EDSS.

Olfactory Assessment

Olfactory function was assessed using the Sniffin' Sticks test (Burghart Medical Technology, Wedel, Germany), a validated psychophysical battery that evaluates three domains: Odor threshold, odor discrimination and odor identification. The three subscores were summed to obtain the TDI composite score.

The test was performed in a well-ventilated, odor-free room, following standardized procedures. Participants were instructed to refrain from consuming food or beverages (other than water), smoking, or using scented products for at least 1 hour before testing. Odorants were presented in felt-tip dispensers. The pen cap was removed for a few seconds, and the tip was held approximately 1-2 cm in front of both nostrils.

Odor threshold was assessed using n-butanol in a 3-alternative forced-choice staircase paradigm. For each trial, one pen contained n-butanol at ascending dilutions (1:2 dilution steps), and two pens contained only solvent. Participants identified the pen with the odorant, and the threshold score was calculated after seven reversals based on the lowest correctly recognized concentration. Odor discrimination was evaluated using 16 triplets of pens, each consisting of two pens with identical odors and one pen with a different odor. Participants were asked to identify the pen with a different odor. Odor identification was assessed using 16 common odorants, each presented with four response options. A minimum 30-second interval was maintained between odor presentations to prevent olfactory desensitization. Each subtest had a maximum score of 16 points, and the total TDI score reflected global olfactory performance (range: 0-48), with higher scores indicating better function (Table 1).

Based on established normative criteria, TDI scores <16.5 were classified as anosmia, scores of 16.5-30.5 as hyposmia, and scores >30.5 as normosmia. Threshold, discrimination, and identification scores of <10, <12, and <11, respectively, were considered abnormal (Table 2).

MRI

MRI was performed using a 1.5 T scanner. MRI abnormalities were evaluated by an experienced neurologist blinded to the olfactory test results. T2-weighted and FLAIR sequences were used to assess hyperintense demyelinating plaques, while T1-weighted sequences were used to assess hypointense lesions. Post-contrast gadolinium-enhanced T1 images were reviewed to determine the presence of active lesions. Plaque localization was recorded across the following regions: Frontal, temporal, parietal, and occipital; brainstem; cerebellum; cervical spinal cord; and periventricular and/or juxtacortical areas.

Statistical Analysis

Statistical analyses were performed using the SPSS 24.0 software package (SPSS Inc., Chicago, IL, USA). Categorical variables were expressed as frequencies and percentages, and continuous variables as means and standard deviations or medians and interquartile ranges, as appropriate. Normality of the distribution was assessed using the Kolmogorov-Smirnov test. Group comparisons were performed using the chi-squared test for categorical variables, and the independent t-test, Mann-Whitney U test, One-Way ANOVA, or Kruskal-Wallis test for continuous variables, depending on distributional assumptions. Correlations between olfactory measures and clinical variables were examined using Pearson's correlation analysis. Values of $p < 0.05$ were considered statistically significant.

Table 1. Classification of olfactory function according to Sniffin' Sticks TDI criteria.

Classification	TDI score range	Interpretation
Anosmia	<16.5	Absence of functional olfaction
Hyposmia	16.5-30.5	Reduced olfactory function
Normosmia	>30.5	Normal olfactory function

TDI: Threshold + discrimination + identification scores (range 0-48).

RESULTS

The patient group comprised 28 patients with MS and one patient with CIS; the control group included 27 healthy participants. There were no significant differences between the groups regarding age, sex distribution, or smoking status (all $p > 0.05$). Demographic and clinical characteristics are presented in Table 3.

Based on the TDI score, one participant in the patient group (3.4%) was classified as anosmic, and 17 (58.6%) were classified as hyposmic. Abnormal threshold, discrimination, and identification scores were observed in 16 (55.2%), 20 (69.0%), and 17 (58.6%) participants, respectively. In the control group, 5 (18.5%) participants were classified as hyposmic based on the TDI score, while abnormal threshold, discrimination, and identification scores were observed in 4 (14.8%), 19 (70.4%), and 8 (29.6%) participants, respectively. Compared with healthy controls, the patient group showed significantly lower odor threshold, identification, and total TDI scores ($p = 0.021$, 0.004, 0.005, respectively) (Table 3).

According to clinical classification, 13 patients (44.8%) had RRMS, 1 patient (3.4%) had CIS, 1 patient (3.4%) had PPMS, and 14 patients (48.3%) had SPMS. 15 patients (51.7%) had a progressive disease course. There were no significant differences in olfactory threshold, discrimination, identification, or total TDI scores across the clinical classifications (all $p > 0.05$). The presence of a progressive disease course did not influence olfactory performance (all $p > 0.05$).

MRI showed periventricular plaques in all 29 participants in the patient group (100%), followed by juxtacortical plaques in 18 (62.1%), frontal plaques in 16 (55.2%), cervical spinal cord plaques in 16 (55.2%), temporal plaques in 12 (41.4%), brainstem plaques in 12 (41.4%), and cerebellar plaques in 5 (17.2%) (Table 4). No significant associations were found between plaque distribution in any region and olfactory scores (all $p > 0.05$).

Correlation analysis showed that, in the patient group, odor discrimination, identification, and total TDI scores were significantly correlated with age ($r = -0.600$, -0.370 and -0.514; $p = 0.001$, 0.048 and 0.004, respectively) and disease duration ($r = -0.592$, -0.409 and -0.579; $p = 0.001$, 0.028 and 0.001, respectively) (Table 5). No

Table 2. Sniffin' Sticks subtest cut-off values and interpretation of abnormal scores.

Subtest	Cut-off score	Interpretation of scores below cut-off
Threshold (T)	<10	Reduced odor detection sensitivity
Discrimination (D)	<12	Impaired differentiation between odor qualities
Identification (I)	<11	Impaired odor recognition

Cut-off values based on validated normative data.

Table 3. Demographic, clinical characteristics and Sniffin' Sticks test scores of the MS and control groups.

	Patient group (n = 29)	Control group (n=27)	p-value
Sex: n (% female)	22 (75.9%)	15 (55.6%)	0.109
Age: years ± SD (min.-max.)	38.86 ± 11.38 (23-65)	29.7 ± 5.94 (19-63)	0.057
Smoking: n (%)	8 (27.6%)	12 (44.4%)	0.188
Disease duration: mean ± SD (years)	11.45 ± 8.29	-	-
EDSS: mean ± SD (min.-max.)	3.5 ± 2.7 (0-8)	-	-
OTS: mean ± SD	9.81 ± 3.49	11.62 ± 1.93	0.021*
ODS: mean ± SD	10.21 ± 2.44	10.7 ± 1.93	0.405
OIS: mean ± SD	9.28 ± 3.2	11.3 ± 1.51	0.004*
TDI: mean ± SD	29.21 ± 6.64	33.51 ± 3.83	0.005*

MS: Multiple sclerosis, SD: Standard deviation, EDSS: Expanded disability status scale, OTS: Odor threshold score, ODS: Odor discrimination score, OIS: Odor identification score, TDI: Threshold, discrimination, and identification.

Table 4. Distribution of demyelinating plaques by anatomical region on MRI in the patient group (n = 29).

Region	n (%)
Periventricular	29 (100%)
Juxtacortical	18 (62.1%)
Frontal	16 (55.2%)
Temporal	12 (41.4%)
Cerebellar	5 (17.2%)
Brainstem	12 (41.4%)
Cervical spinal cord	16 (55.2%)

MRI: Magnetic resonance imaging.

significant correlation was observed between EDSS scores and any olfactory measure (all p > 0.05).

DISCUSSION

Olfactory function is increasingly recognized as a clinically relevant neurobiological marker in MS (6,13). In this study, we demonstrated impaired global olfactory performance in the patient group compared with healthy controls, consistent with previous studies reporting olfactory dysfunction in 30% to 75% of patients with MS (6,22). In our cohort, 58.6% of patients were classified as hyposmic and 3.4% as anosmic based on TDI scores, corresponding to an overall olfactory dysfunction frequency of 62.1%. Odor threshold, identification, and total TDI scores were significantly lower than those of the control group. Odor discrimination did not differ significantly between the patient and control groups, supporting the notion that individual olfactory subdomains may reflect distinct underlying mechanisms.

Consistent with earlier findings, we observed that odor identification, discrimination, and TDI scores were lower with increasing age and longer disease duration (7). This observation supports the hypothesis that olfactory performance reflects cumulative neurodegenerative burden rather than acute inflammatory activity. In contrast, odor threshold impairment has been associated with short-term inflammatory processes, particularly during relapsing-remitting phases (23). Other studies suggest that threshold abnormalities may show sensitivity to inflammatory activity, whereas identification and discrimination deficits more strongly reflect neurodegenerative mechanisms (23-25). Our results align with this view, as abnormal threshold scores were observed in more than half of the patients, suggesting that threshold impairment may indicate ongoing inflammatory activity.

Importantly, we found no significant relationships between olfactory test performance and clinical MS subtype, EDSS score, or MRI plaque distribution. Previous studies have reported mixed results, with some indicating correlations between olfactory dysfunction, higher EDSS scores, and progressive disease course (13,26,27), while others found no such associations (6,17,28,29). Our findings suggest that EDSS scores and plaque counts obtained by conventional MRI may not adequately reflect the functional impact of MS on the olfactory network. Standard imaging techniques often lack the sensitivity to

Table 5. Correlations between age, disease duration, EDSS scores, and olfactory test scores for the patient group.

Sniffin' Sticks test results	Age		Disease duration		EDSS	
	r	p	r	p	r	p
OTS	-0.234	0.221	-0.330	0.080	-0.058	0.766
ODS	-0.600**	0.001	-0.592**	0.001	-0.268	0.160
OIS	-0.370*	0.048	-0.409*	0.028	-0.278	0.144
TDI	-0.514**	0.004	-0.579**	0.001	-0.276	0.148

EDSS: Expanded disability status scale, OTS: Odor threshold score, ODS: Odor discrimination score, OIS: Odor identification score, TDI: Threshold, discrimination, and identification, *: Correlation is significant at the 0.05 level, **: Correlation is significant at the 0.01 level.

detect subtle microstructural and network-level changes that likely contribute to olfactory impairment.

The network-based model of olfactory dysfunction provides an integrative framework for understanding the multisystem nature of olfactory impairment in MS. Consequently, plaque burden detected on conventional MRI may not adequately represent the physiologically relevant network-level damage underlying olfactory dysfunction. Advanced imaging studies have shown significant correlations between olfactory deficits and reduced olfactory bulb volume, frontotemporal cortical thinning, and reduced connectivity across olfactory network pathways (6,7,30-33). In our cohort, plaques were predominantly located in periventricular and juxtacortical regions; however, no significant associations were found between plaque topography and olfactory scores. This finding may be attributed to the limited spatial resolution of standard MRI rather than the absence of true structural involvement.

Taken together, these findings emphasize that olfactory impairment in MS arises from distributed inflammatory and neurodegenerative processes rather than from isolated focal lesions. This highlights the importance of comprehensive olfactory assessment as a clinically feasible and non-invasive approach for detecting subtle network-level alterations that may occur before measurable functional decline.

Among the olfactory subcomponents, threshold impairment appears to reflect ongoing inflammatory activity, whereas identification and discrimination deficits may correspond to cumulative neurodegenerative burden. This pattern has been observed in several recent studies demonstrating that odor threshold fluctuates with relapse-related inflammation, while identification and discrimination scores show progressive reduction with longer disease duration and greater cortical atrophy (13,23-25,34). Evaluating these subcomponents separately, rather than relying solely on the composite TDI score, may provide deeper insights into the transition from inflammation to neurodegeneration in MS and may support the development of olfactory testing as a complementary clinical biomarker.

From a clinical perspective, olfactory assessment represents a simple, inexpensive, and non-invasive tool that can be easily incorporated into routine neurological evaluations. By assessing threshold, discrimination, and identification separately, multidimensional olfactory testing may help distinguish inflammatory activity from cumulative neurodegeneration in people with MS. Integrating olfactory evaluation with neuroimaging and cognitive assessment may enhance early detection of disease progression and identify patients at higher risk for neurodegenerative decline (13). As recent studies have underscored the value of sensory biomarkers in neuroinflammatory disorders, olfactory measures could serve as complementary indicators reflecting both functional and structural aspects of central nervous system involvement (13,25,34). Further longitudinal studies are warranted to determine whether specific olfactory parameters can reliably predict clinical or radiological disease activity, thereby refining their role as complementary biomarkers in MS monitoring.

Study Limitations

This study has several limitations. First, the relatively small sample size and cross-sectional design limit the generalizability of our findings and preclude causal inferences. The small number of patients within individual clinical categories also reduced the statistical power of subgroup comparisons. Therefore, the absence of significant subtype-related differences should be interpreted cautiously. Second, patients experiencing acute relapses or receiving corticosteroid treatment were excluded, limiting our ability to evaluate the direct effects of acute inflammation on olfactory function. Third, only conventional MRI techniques were used, which may not have been sufficiently sensitive to detect subtle microstructural or network-level alterations associated with olfactory impairment. Future longitudinal studies using advanced neuroimaging methods in larger and more diverse MS populations are needed to confirm and extend these findings.

CONCLUSION

Olfactory dysfunction was frequent in patients with MS. Lower discrimination, identification, and total TDI scores were associated with older age and longer disease duration. Threshold impairment may be related to inflammatory processes, whereas lower identification and discrimination scores may reflect cumulative neurodegenerative changes. Comprehensive olfactory assessment is a practical, non-invasive, and inexpensive approach that may complement clinical evaluation in MS. Longitudinal studies are needed to determine whether olfactory measures can be used to monitor neurodegenerative changes or disease progression.

Ethics

Ethics Committee Approval: The study was approved by the Institutional Ethics Committee of Sakarya University, Faculty of Medicine (ethical approval number: E-16214662-050.01.04-28539-97, date: 21.04.2021).

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

Acknowledgments

For transparency, the authors note that an artificial intelligence-assisted language model (ChatGPT, OpenAI) was utilized to support text editing and language correction. This assistance was limited to linguistic refinement; all scientific content, critical analysis, and final editorial decisions were made exclusively by the authors.

Footnotes

Authorship Contributions

Concept: A.A.G., D.K., G.K., Design: A.A.G., D.K., G.K., A.Ö.A., A.A., Data Collection or Processing: A.Ö.A., G.K., A.A., Analysis or Interpretation: A.A.G., D.K., Literature Search: A.A.G., D.K., A.Ö.A., Writing: A.A.G., A.Ö.A.

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

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

REFERENCES

1. Croy I, Nordin S, Hummel T. Olfactory disorders and quality of life—an updated review. Chem Senses. 2014; 39: 185-94.

2. Han SA, Kim JK, Cho DY, Patel ZM, Rhee CS. The olfactory system: basic anatomy and physiology for general otorhinolaryngologists. Clin Exp Otorhinolaryngol. 2023; 16: 308-16.
3. Efendioglu MK, Orhan EK, Sen C, Sonmez S, Orhan KS, Baykan B. Olfactory function assessment of migraine patients by using the Sniffin' Sticks test: a clinical study. Am J Otolaryngol. 2024; 45: 104076.
4. Patino J, Karagas NE, Chandra S, Thakur N, Stimming EF. Olfactory dysfunction in Huntington's disease. J Huntingtons Dis. 2021; 10: 413-22.
5. Dan X, Wechter N, Gray S, Mohanty JG, Croteau DL, Bohr VA. Olfactory dysfunction in aging and neurodegenerative diseases. Ageing Res Rev. 2021; 70: 101416.
6. Lucassen EB, Turel A, Knehans A, Huang X, Eslinger P. Olfactory dysfunction in multiple sclerosis: a scoping review of the literature. Mult Scler Relat Disord. 2016; 6: 1-9.
7. Bsteh G, Berek K, Hegen H, Teuchner B, Auer M, Wurth S, et al. Smelling multiple sclerosis: different qualities of olfactory function reflect either inflammatory activity or neurodegeneration. Mult Scler. 2020; 26: 57-68.
8. Filippi M, Bar-Or A, Piehl F, Preziosa P, Solari A, Vukusic S, et al. Multiple sclerosis. Nat Rev Dis Primers. 2018; 4: 43.
9. Good KP, Tourbier IA, Moberg PJ, Cuzzocreo JL, Geckle RJ, Yousem DM, et al. Unilateral olfactory sensitivity in multiple sclerosis. Physiol Behav. 2017; 168: 24-30.
10. DeLuca GC, Joseph A, George J, Yates RL, Hamard M, Hofer M, et al. Olfactory pathology in central nervous system demyelinating diseases. Brain Pathol. 2015; 25: 543-51.
11. Schmidt FA, Maas MB, Geran R, Schmidt C, Kunte H, Ruprecht K, et al. Olfactory dysfunction in patients with primary progressive MS. Neurol Neuroimmunol Neuroinflamm. 2017; 4: e369.
12. Doty RL, Li C, Mannon LJ, Yousem DM. Olfactory dysfunction in multiple sclerosis: relation to longitudinal changes in plaque numbers in central olfactory structures. Neurology. 1999; 53: 880-2.
13. Printza A, Boziki M, Valsamidis C, Bakirtzis C, Constantinidis J, Grigoriadis N, et al. Smell as a disease marker in multiple sclerosis. J Clin Med. 2022; 11: 5215.
14. Kulason S, Ratnanather JT, Miller MI, Kamath V, Hua J, Yang K, et al. A comparative neuroimaging perspective of olfaction and higher-order olfactory processing: on health and disease. Semin Cell Dev Biol. 2022; 129: 22-30.
15. Almasi M, Sahraian MA, Akhoundi FH, Ezzati HR, Rohani M. The factors associated with olfactory dysfunction in patients with multiple sclerosis. Basic Clin Neurosci. 2021; 12: 89-94.
16. Uecker FC, Olze H, Kunte H, Gerz C, Göktas O, Harms L, et al. Longitudinal testing of olfactory and gustatory function in patients with multiple sclerosis. PLoS One. 2017; 12: e0170492.
17. Batur Caglayan HZ, Irkec C, Nazliel B, Akyol Gurses A, Capraz I. Olfactory functioning in early multiple sclerosis: Sniffin' Sticks test study. Neuropsychiatr Dis Treat. 2016; 12: 2143-7.
18. Kandemir S, Muluk NB, Melikoglu B, Dag E, Inal M, Sarin O. Smell functions in patients with multiple sclerosis: a prospective case-control study. B-ENT. 2016; 12: 323-31.
19. Atalar AC, Erdal Y, Tekin B, Yildiz M, Akdogan O, Emre U. Olfactory dysfunction in multiple sclerosis: relation with magnetic resonance imaging findings and disease parameters. Mult Scler Relat Disord. 2018; 21: 92-6.
20. Öniz A, Erdoğan İ, İkiz AÖ, Evirgen N, Özgören M. The modified Sniffin' Sticks test in Turkish population based on odor familiarity survey. J Neurol Sci Turk. 2013; 30: 270-80.
21. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 2018; 17: 162-73.
22. Joseph A, DeLuca GC. Back on the scent: the olfactory system in CNS demyelinating diseases. J Neurol Neurosurg Psychiatry. 2016; 87: 1146-54.
23. Lutterotti A, Vedovello M, Reindl M, Ehling R, Di Pauli F, Kuenz B, et al. Olfactory threshold is impaired in early, active multiple sclerosis. Mult Scler. 2011; 17: 964-9.
24. Carotenuto A, Costabile T, Moccia M, Falco F, Scala MR, Russo CV, et al. Olfactory function and cognition in relapsing-remitting and secondary-progressive multiple sclerosis. Mult Scler Relat Disord. 2019; 27: 1-6.
25. Bsteh G, Hegen H, Ladstätter F, Berek K, Amprosi M, Wurth S, et al. Change of olfactory function as a marker of inflammatory activity and disability progression in MS. Mult Scler. 2019; 25: 267-74.
26. Mirmosayyeb O, Ebrahimi N, Barzegar M, Afshari-Safavi A, Bagherieh S, Shaygannejad V. Olfactory dysfunction in patients with multiple sclerosis; a systematic review and meta-analysis. PLoS One. 2022; 17: e0266492.
27. Zorzon M, Ukmar M, Bragadin LM, Zanier F, Antonello RM, Cazzato G, et al. Olfactory dysfunction and extent of white matter abnormalities in multiple sclerosis: a clinical and MR study. Mult Scler. 2000; 6: 386-90.
28. Jordy SS, Starzewski Junior A, Macedo FAB, Manica GR, Tilbery CP, Carabetta EG. Olfactory alterations in patients with multiple sclerosis. Arq Neuropsiquiatr. 2016; 74: 697-700.
29. Todd LL, Sivakumar R, Lynch SG, Diebolt JH, White J, Villwock JA. Longitudinal olfactory patterns in multiple sclerosis: a scoping review and implication for use in management of disease. Int J MS Care. 2023; 25: 131-6.
30. Goektas O, Schmidt F, Bohner G, Erb K, Lüdemann L, Dahlslett B, et al. Olfactory bulb volume and olfactory function in patients with multiple sclerosis. Rhinology. 2011; 49: 221-6.
31. Holinski F, Schmidt F, Dahlslett SB, Harms L, Bohner G, Olze H. MRI study: objective olfactory function and CNS pathologies in patients with multiple sclerosis. Eur Neurol. 2014; 72: 157-62.
32. Yaldizli Ö, Penner IK, Yonekawa T, Naegelin Y, Kuhle J, Pardini M, et al. The association between olfactory bulb volume, cognitive dysfunction, physical disability and depression in multiple sclerosis. Eur J Neurol. 2016; 23: 510-9.
33. Ciurleo R, De Salvo S, Caminiti F, Militi A, Bonanno L. Olfactory evoked potentials and brain MRI outcomes in multiple sclerosis patients: a cross-sectional study. J Clin Med. 2025; 14: 141.
34. Doty RL. Olfactory dysfunction in neurodegenerative diseases: is there a common pathological substrate? Lancet Neurol. 2017; 16: 478-88.

Original Investigation | Özgün Araştırma

Gazi Med J 2026;37(4):467-474

DOI: http://dx.doi.org/10.12996/gmj.2026.4664

Effects of Oral *Channa striata* and Topical *Centella asiatica* on Serum IL-6 and Wound Tissue TNF-α in a Diabetic Incisional Wound Model in Wistar Rats

Wistar Sıçanlarında Diyabetik İnsizyonel Yara Modelinde Oral *Channa striata* ve Topikal *Centella asiatica*'nın Serum IL-6 ve Yara Dokusu TNF-α Üzerindeki Etkileri

Dimas Irfan Nabih[1], Eko Setiawan[2,3], Joko Wahyu Wibowo[1]

[1]Department of Postgraduate Biomedical Science, Faculty of Medicine, Universitas Islam Sultan Agung, Semarang, Indonesia
[2]Department of Doctoral Biomedical Science, Faculty of Medicine, Universitas Islam Sultan Agung, Semarang, Indonesia
[3]Department of Surgery, Faculty of Medicine, Universitas Islam Sultan Agung, Semarang, Indonesia

ABSTRACT

Objective: To determine the effects of topical 10% *Centella asiatica* gel, oral *Channa striata* extract, and their combination on systemic interleukin (IL)-6 and wound tissue tumor necrosis factor(TNF)-α levels in a Wistar rat model of diabetic incision wound.

Methods: This *in vivo* experimental study used a post-test-only control group design. Diabetes was induced with nicotinamide followed by streptozotocin. Male Wistar rats aged 10 weeks and weighing 180-200 g were randomized into five groups (n = 5/group): Non-diabetic wound control, diabetic wound plus vehicle, diabetic wound plus oral *Channa striata* extract, diabetic wound plus topical 10% *Centella asiatica* gel, and diabetic wound plus combination therapy. Serum IL-6 and wound tissue TNF-α were measured by sandwich ELISA at study termination.

Results: Serum IL-6 differed significantly among groups (One-Way ANOVA, p = 0.008). Compared with the diabetic vehicle group, serum IL-6 was significantly lower in the oral Channa striata group (p = 0.004) and the combination group (p = 0.043). The oral *Channa striata* group did not differ significantly from the combination group (Tukey HSD, p = 0.796). Wound tissue TNF-α did not differ significantly among groups (Kruskal-Wallis, p = 0.100).

Conclusion: Oral *Channa striata*, either alone or combined with topical *Centella asiatica* gel, was associated with lower serum IL-6, whereas wound tissue TNF-α was not significantly changed. The combination did not show an additive serum IL-6 reduction beyond oral monotherapy in this model.

ÖZ

Amaç: Diyabetik insizyonel yara oluşturulan Wistar sıçanı modelinde, topikal %10 *Centella asiatica* jeli, oral *Channa striata* ekstresi ve bu iki uygulamanın kombinasyonunun sistemik interlökin (IL)-6 ve yara dokusu tümör nekroz faktörü (TNF)-α düzeyleri üzerindeki etkilerini belirlemek.

Yöntemler: Bu *in vivo* deneysel çalışmada, yalnızca son test ölçümünün yapıldığı kontrol gruplu deneysel tasarım kullanıldı. Diyabet, nikotinamid uygulamasını takiben streptozotosin verilerek indüklendi. On haftalık, 180-200 g ağırlığındaki erkek Wistar sıçanları, her grupta beş hayvan olacak şekilde (n = 5/grup) rastgele beş gruba ayrıldı: Diyabetik olmayan yara kontrol grubu, taşıyıcı uygulanan diyabetik yara grubu, oral *Channa striata* ekstresi uygulanan diyabetik yara grubu, topikal %10 *Centella asiatica* jeli uygulanan diyabetik yara grubu ve kombinasyon tedavisi uygulanan diyabetik yara grubu. Çalışma sonunda serum IL-6 ve yara dokusu TNF-α düzeyleri sandviç ELISA yöntemiyle ölçüldü.

Bulgular: Serum IL-6 düzeyleri gruplar arasında istatistiksel olarak anlamlı farklılık gösterdi (tek-yönlü varyans analizi, p = 0,008). Diyabetik taşıyıcı grubuyla karşılaştırıldığında, serum IL-6 düzeyleri oral *Channa striata* ekstresi uygulanan grupta (p = 0,004) ve kombinasyon tedavisi uygulanan grupta (p = 0,043) anlamlı olarak daha düşüktü. Oral *Channa striata* ekstresi uygulanan grup ile kombinasyon tedavisi uygulanan grup arasında istatistiksel olarak anlamlı bir farklılık saptanmadı (Tukey HSD testi, p = 0,796). Yara dokusu TNF-α düzeyleri ise gruplar arasında

Cite this article as: Nabih DI, Setiawan E, Wibowo JW. Effects of oral *Channa striata* and topical *Centella asiatica* on serum IL-6 and wound tissue TNF-α in a diabetic incisional wound model in Wistar rats. Gazi Med J. 2026;37(4):467-474

Address for Correspondence/Yazışma Adresi: Eko Setiawan, Department of Doctoral Biomedical Science; Department of Surgery, Faculty of Medicine, Universitas Islam Sultan Agung, Semarang, Indonesia
E-mail / E-posta: drekosetiawan@unissula.ac.id
ORCID ID: orcid.org/0000-0002-7194-2359

Received/Geliş Tarihi: 19.03.2026
Accepted/Kabul Tarihi: 27.07.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

Keywords: *Centella asiatica*, *Channa striata*, diabetic wound, interleukin-6, tumor necrosis factor-alpha, diabetes mellitus

ÖZ

istatistiksel olarak anlamlı farklılık göstermedi (Kruskal-Wallis testi, p = 0,100).

Sonuç: Oral *Channa striata* ekstresi, tek başına veya topikal *Centella asiatica* jeli ile birlikte uygulandığında daha düşük serum IL-6 düzeyleriyle ilişkili bulundu; buna karşılık yara dokusu TNF-α düzeylerinde istatistiksel olarak anlamlı bir değişiklik saptanmadı. Bu modelde, kombinasyon tedavisinin serum IL-6 düzeylerini oral monoterapinin sağladığı azalmanın ötesinde düşürdüğüne ilişkin bir bulgu elde edilmedi.

Anahtar Sözcükler: *Centella asiatica*, *Channa striata*, diyabetik yara, interlökin-6, tümör nekroz faktörü-alfa, diabetes mellitus

INTRODUCTION

Diabetes mellitus remains a major global health burden, with chronic wounds among its most serious complications. Persistent hyperglycemia disrupts normal wound healing by causing vascular dysfunction, impaired cellular responses, and prolonged inflammation. Diabetic wounds are characterized by excessive inflammatory activity, including dysregulation of pro-inflammatory cytokines such as tumor necrosis factor-alpha (TNF-α) and interleukin-6 (IL-6). This sustained inflammatory state increases proteolytic activity, impairs extracellular matrix formation, and delays progression to the proliferative and remodeling phases, resulting in poor wound healing and adverse clinical outcomes (1-3).

Recent evidence supports the use of botanical and nutraceutical adjuvants to optimize wound healing in diabetes. *Centella asiatica* has been widely investigated for topical wound care; contemporary reviews and experimental studies highlight triterpenoid-rich constituents, including asiaticoside and madecassoside, that may promote fibroblast activity, collagen synthesis, re-epithelialization, angiogenesis, and inflammatory regulation (4-7). In parallel, *Channa striata* supplementation has gained attention as a systemic nutritional therapy rich in albumin and essential amino acids; clinical and experimental reports suggest acceleration across inflammatory, proliferative, and maturation phases of wound repair, including in diabetic wound contexts (8-10). A recent systematic review and meta-analysis also indicates that topical and oral herbal products can support diabetic foot ulcer healing, strengthening the rationale for integrative approaches (11).

Despite this progress, the current state of the art remains fragmented. *Centella asiatica* is commonly evaluated as topical monotherapy or in topical combinations, whereas *Channa striata* is frequently studied as a standalone oral intervention with limited concurrent assessment of inflammatory biomarkers (6-8,10). Consequently, there is limited mechanistic evidence on whether a dual-route strategy, consisting of local topical therapy plus systemic oral support, can differentially affect systemic and local inflammatory biomarkers, including serum IL-6 and wound tissue TNF-α in diabetic incisional wounds. Therefore, this study aimed to evaluate the effects of topical *Centella asiatica* gel, oral *Channa striata* extract, and their combination on serum IL-6 and wound tissue TNF-α levels in a streptozotocin-nicotinamide-induced diabetic Wistar rat incisional wound model.

MATERIALS AND METHODS

Study Design and Animals

This *in vivo* study used a post-test-only control group design and was conducted at the IBL Laboratory, Universitas Islam Sultan Agung, Semarang, Indonesia, from November 2025 to January 2026. Wistar rats weighing 180-200 g were housed with a 12-hour light/dark cycle and chow and water ad libitum. After 7 days of acclimatization, animals were allocated into five groups: G1, non-diabetic wound with topical base gel and oral vehicle; G2, diabetic wound with topical base gel and oral vehicle; G3, diabetic wound with oral *Channa striata* extract and topical base gel; G4, diabetic wound with topical 10% *Centella asiatica* gel and oral vehicle; and G5, diabetic wound with combined oral *Channa striata* extract and topical 10% *Centella asiatica* gel.

After acclimatization, each rat was assigned a unique identification code. Rats assigned to the non-diabetic wound control group were allocated to G1. Rats intended for the diabetic groups underwent streptozotocin-nicotinamide induction, and random blood glucose was measured 72 hours later. Only rats with random blood glucose levels ≥200 mg/dL were considered eligible for diabetic group allocation. Eligible diabetic rats were then randomly allocated to G2, G3, G4, or G5 using a computer-generated random number list, with a final target of five rats per group.

As illustrated in Figure 1, several subjects could not be included in the final analysis during the study period. Four subjects did not meet the inclusion criteria based on post-induction random blood glucose levels, as their values did not reach the threshold specified in the protocol, and were therefore categorized as exclusions. In addition, one subject in Group 2 died during the study period and was categorized as a dropout. Therefore, the statistical analysis and reporting of results were conducted using an effective sample size of five rats in each treatment group.

The investigator responsible for outcome data collection, including ELISA-related outcome measurement and data recording, was blinded to group allocation to minimize assessment bias. Treatment administration was performed according to the assigned intervention groups, whereas outcome measurements were conducted using coded animal and sample identifiers without revealing group information to the data collector.

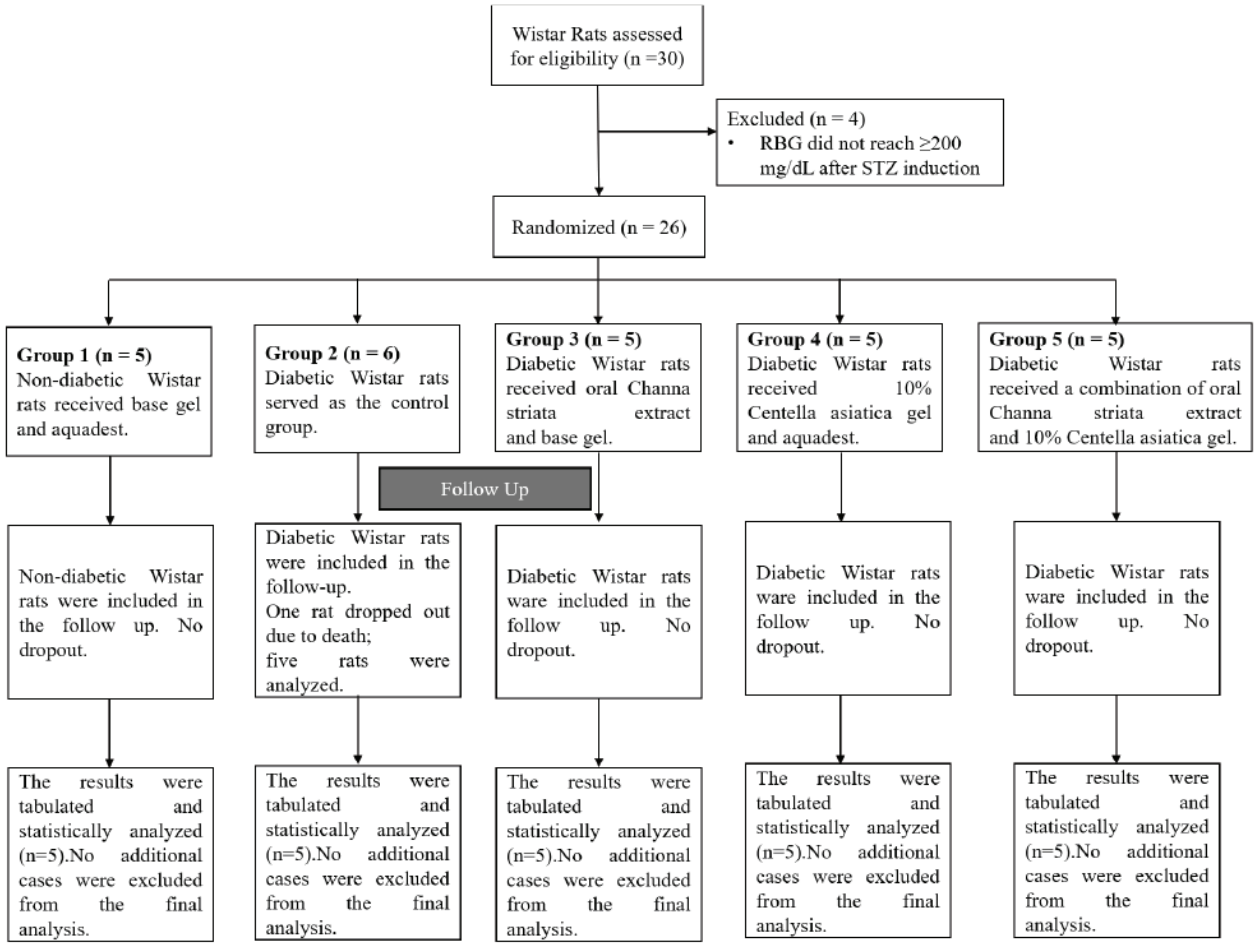

Figure 1. Animal flow diagram of the study. Thirty animals were initially prepared. Four animals were excluded because post-induction random blood glucose levels did not reach the diabetic threshold specified in the protocol. One animal in G2 died during the study period and was categorized as a dropout. The final analysis included 25 animals, with five rats in each group.

Sample size was estimated using the Federer formula, (t-1)(n-1) ≥15, as a minimum sample-size approach for exploratory animal experiments involving multiple groups. This calculation yielded a minimum of five rats per group. A formal a priori power analysis based on predefined effect size, alpha level, and target power was not performed. Therefore, this study should be interpreted as an exploratory preclinical study, particularly for outcomes with potentially modest between-group differences.

Ethical Approval

Procedures were approved by the Bioethics Committee of Sultan Agung Islamic University, Semarang, Indonesia (no: 671/XII/2025/ Komisi Bioetik, date: 15.12.2025). Patient consent was not applicable because this experimental study involved animals and did not include human participants.

Diabetes Induction and Wound Model

After overnight fasting, diabetes was induced by intraperitoneal nicotinamide at 120 mg/kg, followed 15 minutes later by streptozotocin at 65 mg/kg. Streptozotocin was freshly prepared immediately before injection by dissolving it in cold sodium citrate buffer, pH 4.5. The streptozotocin solution was protected from light during preparation and administered shortly after dissolution to minimize degradation. The injection volume was calculated individually according to each rat's body weight and the prepared streptozotocin concentration to deliver the intended dose. Random blood glucose was measured from the tail vein 72 hours later, and rats with levels ≥200 mg/dL were considered diabetic. For wound creation, animals were anesthetized with ketamine (80 mg/kg) and xylazine (10 mg/kg). The dorsal area was shaved, disinfected with povidone-iodine, and a 2-cm linear incision reaching the subcutaneous layer was created on the dorsal skin. The wound was left open.

An incisional wound model was selected because the primary objective of this study was to evaluate inflammatory biomarker responses in a standardized local wound environment, rather than to quantify macroscopic wound contraction or closure rate. Incisional wounds provide a reproducible linear injury with controlled length and depth, allowing wound-edge tissue sampling for local cytokine assessment. In contrast, excisional wound models are more commonly used when the main outcomes are wound area reduction,

contraction, granulation tissue formation, and closure dynamics. Therefore, the incisional model was considered appropriate for this exploratory biomarker-focused study, while the absence of wound contraction and closure-rate measurements was acknowledged as a limitation (12,13).

Interventions

Interventions were administered for 7 consecutive days. G1 and G2 received topical base gel and oral vehicle. G3 received oral *Channa striata* extract (100 mg/100 g body weight/day) with topical base gel. G4 received topical 10% *Centella asiatica* gel (200 μg/cm^2) with oral vehicle. G5 received both interventions at the same doses. The *Channa striata* extract was reconstituted before use, and the *Centella asiatica* gel was prepared in a carbopol-based formulation containing 10% ethanolic extract.

Before formulation, the *Centella asiatica* ethanolic extract underwent qualitative triterpenoid screening and total flavonoid determination using a quercetin standard curve. Asiaticoside and madecassoside were not quantified. The *Channa striata* extract was not analyzed for albumin or total protein content.

Doses were selected as fixed exploratory preclinical doses based on prior wound-healing evidence for each component, rather than on formal dose-response optimization. The oral *Channa striata* dose of 100 mg/100 g body weight/day was selected to provide systemic nutritional support, because previous oral *Channa striata* studies reported improved wound-repair parameters, including tensile strength, fibroblast count, and hydroxyproline content, and systematic evidence suggests that *Channa striata* may exert wound-healing and anti-inflammatory effects through albumin, amino acids, fatty acids, and other bioactive constituents (8-10). The topical 10% *Centella asiatica* gel dose at 200 μg/cm^2 was selected to provide local wound-bed exposure, because topical *Centella asiatica* preparations and triterpenoid-rich formulations have been associated with improved wound repair through anti-inflammatory, antioxidant, fibroblast-stimulating, collagen-regulating, angiogenic, and re-epithelialization mechanisms (4-7,11). Because this study evaluated a dual-route combination rather than a dose-response model, one fixed dose of each intervention was used to assess preliminary biological plausibility.

Outcome Measures and Sample Collection

Random blood glucose was reassessed on day 8. At the end of treatment, blood was collected for IL-6, centrifuged, and stored until analysis. Wound-edge tissue was excised for TNF-α, rinsed with phosphate-buffered saline, and homogenized in lysis buffer.

Serum IL-6 was selected to represent systemic inflammatory activity, whereas wound tissue TNF-α was selected to represent the local inflammatory response at the wound site. Because these biomarkers were measured in different biological compartments, their responses were interpreted as complementary rather than directly equivalent indicators of inflammation.

ELISA Assay

Serum IL-6 and wound tissue TNF-α were quantified using commercially available sandwich ELISA kits specific for rat samples. Serum IL-6 was measured using a Rat IL-6 Quantikine ELISA Kit (R&D Systems, USA; catalog no. R6000B; species reactivity: rat), whereas wound tissue TNF-α was measured using a Rat TNF-α Quantikine ELISA Kit (R&D Systems, USA; catalog no. RTA00; species reactivity: rat). For wound tissue TNF-α measurement, wound-edge tissue was weighed before homogenization, and tissue samples were homogenized using a standardized tissue-to-lysis buffer ratio. Specifically, 1 mL of fresh lysis buffer was added for every 200 mg of wound tissue. The homogenate was centrifuged, and the supernatant was used for ELISA analysis. TNF-α concentrations were calculated from the standard curve and expressed as pg/mL of tissue homogenate. The values were not normalized to total protein concentration because total protein quantification was not performed. Serum IL-6 concentrations were expressed as pg/mL. Absorbance was read at 450 nm with correction at 570/630 nm, and concentrations were calculated from standard curves according to the manufacturers' instructions.

Statistical Analysis

Data are presented as mean ± standard deviation. Normality was assessed using the Shapiro-Wilk test and homogeneity using Levene's test. Pre-treatment blood glucose levels were compared across groups using Welch's ANOVA followed by the Games-Howell post-hoc test. Post-treatment blood glucose levels and serum IL-6 levels were analyzed using One-Way ANOVA followed by Tukey's HSD post-hoc test. Within-group blood glucose changes were analyzed using the Wilcoxon signed-rank test. Wound tissue TNF-α levels were analyzed using the Kruskal-Wallis test because one group violated the normality assumption and the sample size was small, despite homogeneous variance based on Levene's test. This conservative non-parametric approach was selected to reduce reliance on normal-distribution assumptions. A p-value < 0.05 was considered statistically significant. Analyses were performed using IBM SPSS Statistics version 26.0 for Windows.

RESULTS

Phytochemical screening of the *Centella asiatica* ethanolic extract showed a positive triterpenoid result, indicated by the formation of an orange ring after the addition of acetic anhydride and H_2SO_4. Total flavonoid analysis using a quercetin standard curve (10–100 ppm; $R^2 = 0.9988$) revealed a mean absorbance of 0.64 and a total flavonoid content of 0.4006 mgQE/g extract.

As shown in Table 1, pre-treatment random blood glucose levels confirmed clear glycemic separation between the non-diabetic control group and the diabetic groups. G1 showed normoglycemic values, with a mean RBG of 104.2 ± 11.78 mg/dL and a range of 93-122 mg/dL. In contrast, the diabetic groups showed substantially higher pre-treatment RBG values, ranging from 344.6 ± 27.36 mg/dL in G5 to 396.2 ± 61.27 mg/dL in G4, confirming successful induction of diabetes before treatment allocation. After treatment, G1 remained normoglycemic, with a mean RBG of 78.4 ± 21.27 mg/dL. The diabetic groups continued to show higher RBG levels than G1, although descriptive variation was observed across groups. Among the diabetic groups, the highest post-treatment mean RBG was observed in G2 (355.0 ± 25.24 mg/dL), whereas the lowest was observed in G5 (290.6 ± 49.55 mg/dL).

As shown in Table 2 and Figure 2, serum IL-6 levels varied across groups, with the highest mean observed in G2 (6.478 ± 1.256 pg/

mL) and the lowest mean observed in G3 (4.064 ± 0.509 pg/mL). The combination group (G5) showed a lower IL-6 mean (4.708 ± 0.609) than G2. Assumption testing supported parametric comparison, with Shapiro-Wilk p > 0.05 for all groups and Levene's test p = 0.512. One-Way ANOVA confirmed a statistically significant difference in serum IL-6 among groups (F = 4.703; p = 0.008), with a large effect size (eta-squared = 0.485).

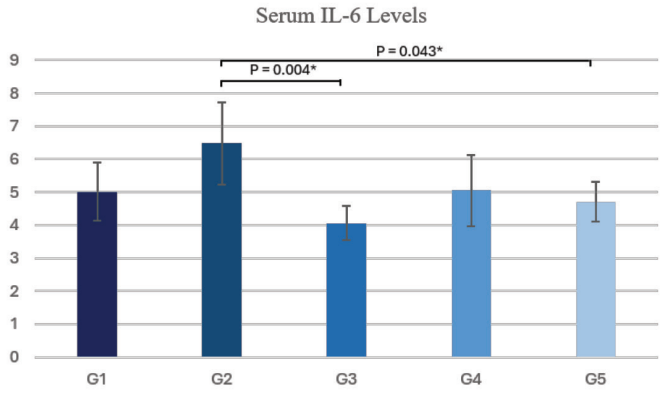

Figure 2. Serum IL-6 concentrations across experimental groups. Data are presented as mean ± SD (pg/mL), n = 5 per group. Statistical analysis was performed using One-Way ANOVA followed by Tukey HSD post-hoc test. *p < 0.05 for the indicated pairwise comparison.

Data are presented as mean ± SD. The asterisk (*) indicates statistically significant post-hoc differences between the indicated groups (p < 0.05). SD: Standard deviation, IL: Interluekin.

Post-hoc analysis using Tukey HSD showed that the statistically significant differences in serum IL-6 were primarily driven by comparisons involving G2. Serum IL-6 was significantly higher in G2 than in G3 (mean difference = 2.414, p = 0.004) and G5 (mean difference = 1.770, p = 0.043). Other relevant post-hoc comparisons were not statistically significant, including G1 vs. G2 (p = 0.121), G2 vs. G4 (p = 0.139), G3 vs. G4 (p = 0.445), G3 vs. G5 (p = 0.796), and G4 vs. G5 (p = 0.973). These findings indicate that the combination group did not differ significantly from either oral *Channa striata* monotherapy or topical *Centella asiatica* monotherapy and did not demonstrate an additional systemic IL-6 reduction beyond oral *Channa striata* alone.

Table 3 and Figure 3 summarize wound tissue TNF-α concentrations expressed as pg/mL of tissue homogenate, together with assumption testing and the between-group comparison. Mean TNF-α was 230.82 ± 16.41 in G1, 234.12 ± 19.71 in G2, 245.44 ± 4.26 in G3, 239.91 ± 15.11 in G4, and 252.83 ± 5.83 in G5. Shapiro-Wilk testing identified a deviation from normality in G2 (p = 0.042), whereas the remaining groups satisfied the normality assumption (p > 0.05). Variance homogeneity was supported by Levene's test (p = 0.112). Although variance homogeneity was satisfied, the Kruskal-Wallis test was selected because one group violated the normality assumption and the sample size was small. This conservative non-parametric approach was used to reduce reliance on normal-distribution

Table 1. Pre - and post-treatment random blood glucose levels across experimental groups.

Group	n	Pre-treatment RBG, mean ± SD (mg/dL) (min-max)	Post-treatment RBG, mean ± SD (mg/dL) (min-max)
G1	5	104.2 ± 11.78 (93-122)	78.4 ± 21.27 (58-112)
G2	5	381.6 ± 50.76 (329-445)	355.0 ± 25.24 (326-392)
G3	5	393.8 ± 31.29 (361-425)	327.0 ± 35.27 (292-364)
G4	5	396.2 ± 61.27 (314-465)	336.0 ± 30.09 (298-374)
G5	5	344.6 ± 27.36 (310-382)	290.6 ± 49.55 (216-339)

RBG: Random blood glucose, SD: Standard deviation, G1: Non-diabetic wound control, G2: Diabetic wound plus vehicle, G3: Diabetic wound plus oral *Channa striata* extract, G4: Diabetic wound plus topical 10% *Centella asiatica* gel, G5: Diabetic wound plus combination therapy.

Table 2. Serum IL-6 concentrations (pg/mL): Descriptive statistics, assumption tests, and One-Way ANOVA.

Parameter	G1	G2	G3	G4	G5
Mean ± SD (pg/mL)	5.012 ± 0.883	6.478 ± 1.256	4.064 ± 0.509	5.056 ± 1.081	4.708 ± 0.609
Shapiro-Wilk p-value	0.975	0.357	0.806	0.893	0.954
Levene p-value (based on mean)	0.512				
One-Way ANOVA p-value	0.008*				

Shapiro-Wilk p > 0.05 indicates normal distribution. Levene p>0.05 indicates homogeneous variances, *: Indicates p < 0.05 (statistically significant), SD: Standard deviation, G1: Non-diabetic wound control, G2: Diabetic wound plus vehicle, G3: Diabetic wound plus oral *Channa striata* extract, G4: Diabetic wound plus topical 10% *Centella asiatica* gel, G5: Diabetic wound plus combination therapy, IL: Interleukin.

assumptions. The Kruskal-Wallis test did not demonstrate a statistically significant difference in wound tissue TNF-α across groups (χ^2 = 7.769, df = 4, p = 0.100).

Individual animal-level values for pre-treatment random blood glucose, post-treatment random blood glucose, serum IL-6, and wound tissue TNF-α are provided in Supplementary Table S1.

DISCUSSION

This study showed a clearer effect on systemic IL-6 than on wound tissue TNF-α. Oral *Channa striata* (G3) and the combined intervention of topical *Centella asiatica* plus oral *Channa striata* (G5) were associated with significantly lower serum IL-6 compared with the diabetic vehicle group (G2), whereas wound tissue TNF-α did not differ significantly among groups. This pattern suggests that the tested interventions modulated systemic inflammatory tone more detectably than local TNF-α expression at the selected endpoint. Because blood glucose changes among diabetic groups were not significant, the IL-6 reduction was more likely related to inflammatory modulation than to short-term glycemic improvement.

The key negative finding of this study is the absence of a clear combination advantage. Although both oral *Channa striata* monotherapy and the combination treatment showed lower serum IL-6 than the diabetic vehicle group, the combination group was not significantly different from oral *Channa striata* alone (G3 vs. G5, p = 0.796). Therefore, the present data do not support superiority of the combined intervention over oral *Channa striata* monotherapy, and the findings should not be interpreted as evidence of additive or synergistic anti-inflammatory effects.

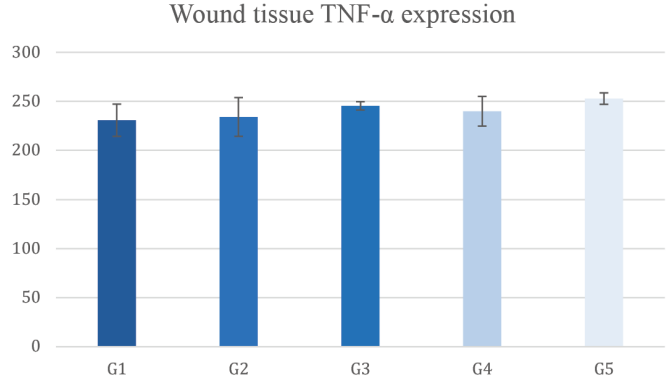

Figure 3. Wound tissue TNF-α concentrations across experimental groups. Data are presented as mean ± SD (pg/mL tissue homogenate), n = 5 per group. Statistical analysis was performed using the Kruskal-Wallis test. No statistically significant difference was observed among groups (p = 0.100).

SD: Standard deviation, TNF: Tissue tumor necrosis factor.

The diabetic vehicle group showed the highest serum IL-6, which was consistent with the inflammatory biology of diabetic wounds. Hyperglycemia can sustain oxidative stress, immune-cell dysfunction, macrophage dysregulation, and delayed inflammatory resolution, thereby maintaining a pro-inflammatory cytokine milieu that includes IL-6 and TNF-α (14-17). However, the difference in serum IL-6 between the non-diabetic wound control group (G1) and the diabetic vehicle group (G2) was not statistically significant. This finding does not invalidate the diabetic wound model, because successful diabetes induction was supported by clear post-induction hyperglycemia and sustained elevation of random blood glucose in the diabetic groups (18). Nevertheless, it indicates that serum IL-6 alone may not function as a robust systemic inflammatory discriminator between non-diabetic and diabetic wounded rats in this experimental setting, since circulating IL-6 is generally interpreted as a marker of peripheral or systemic inflammatory activity (19). Several factors may explain this pattern, including the acute incisional wound stimulus in both G1 and G2, the short observation window, biological variability, and the limited sample size, as wound healing involves dynamic and phase-dependent inflammatory responses mediated by immune cells and pro-inflammatory cytokines (20). Therefore, the IL-6 results should be interpreted primarily as treatment-associated systemic inflammatory modulation within this model, rather than as definitive evidence that serum IL-6 independently distinguished diabetic from non-diabetic wound inflammation. In this context, the lower IL-6 levels observed in G3 and G5 should be interpreted as attenuation of systemic inflammatory activity, not as complete suppression of the inflammatory phase required for tissue repair.

The most important interpretive point is that the serum IL-6 effect appeared to be driven primarily by the oral *Channa striata* component. This interpretation was supported by the post-hoc pattern: G3 had the lowest mean IL-6, G5 was also lower than G2, and G3 did not differ significantly from G5 (p = 0.796). Oral *Channa striata* may influence systemic inflammatory status through albumin-rich nutritional support, amino acid availability, and bioactive lipid or protein-related components that participate in tissue repair and immune regulation (8-10). Therefore, the dual-route design remains theoretically plausible, but the systemic cytokine signal observed in this study was not additive when topical *Centella asiatica* was combined with oral *Channa striata*.

Table 3. Wound tissue TNF-**α** concentrations (pg/mL tissue homogenate): Descriptive statistics and assumption tests (Shapiro-Wilk, Levene) with Kruskal-Wallis.

Parameter	G1	G2	G3	G4	G5
Mean ± SD	230.82 ± 16.41	234.12 ± 19.71	245.44 ± 4.26	239.91 ± 15.11	252.83 ± 5.83
Shapiro-Wilk p-value	0.878	0.042	0.901	0.267	0.440
Levene p-value (based on mean)	0.112				
Kruskal-Wallis p-value	0.100				

Shapiro-Wilk p > 0.05 indicates normal distribution. Levene p > 0.05 indicates homogeneous variances, SD: Standard deviation, G1: Non-diabetic wound control, G2: Diabetic wound plus vehicle, G3: Diabetic wound plus oral *Channa striata* extract, G4: Diabetic wound plus topical 10% *Centella asiatica* gel, G5: Diabetic wound plus combination therapy, TNF: Tissue tumor necrosis factor.

The absence of additive IL-6 reduction in the combination group may have several biological explanations. First, oral *Channa striata* may have produced a near-maximal systemic IL-6 response within the short treatment window, leaving limited measurable room for further reduction by the topical component. Second, topical *Centella asiatica* may act predominantly at the wound microenvironment rather than in the circulation, through effects on fibroblast activity, extracellular matrix formation, re-epithelialization, angiogenesis, oxidative stress, and local inflammatory signaling (4-7). Third, a single serum cytokine at one endpoint may not capture local reparative benefits of topical therapy. Thus, the present findings do not negate the rationale for combination therapy, but they indicate that systemic IL-6 alone is insufficient to prove an added contribution of topical *Centella asiatica*.

The TNF-α findings require careful interpretation. Although mean wound tissue TNF-α was numerically highest in G5, the overall between-group difference was not statistically significant ($p = 0.100$). This directional pattern warrants discussion because the combination group showed a higher mean TNF-α value than both the non-diabetic wound control group and the diabetic vehicle group. One possible interpretation is that the combined intervention may have been associated with transiently greater local inflammatory signaling at day 7, which could reflect ongoing wound-edge remodeling rather than treatment failure (21,22). However, this interpretation remains speculative because wound inflammation is dynamic and phase-dependent, particularly in diabetic wounds, and the present study did not include serial cytokine measurements, histological assessment, wound-closure analysis, collagen deposition, or angiogenesis endpoints to distinguish adaptive inflammatory activity from persistent maladaptive inflammation (14,21,22). We therefore interpret this pattern primarily as biological variability and limited statistical resolution rather than evidence that the combination therapy increased local inflammation.

Serum IL-6 and wound tissue TNF-α were measured in different biological compartments; therefore, serum IL-6 was interpreted as a marker of systemic or peripheral inflammatory activity, whereas wound tissue TNF-α was interpreted as a marker of local wound-site inflammation. These outcomes were considered complementary and were not directly compared as equivalent inflammatory readouts (19,21,22).

TNF-α expression in wounds is dynamic and phase-dependent; at approximately one week after injury, local inflammatory signaling may still coexist with active repair processes. Consequently, local TNF-α values should not be assumed to parallel serum IL-6, and a single endpoint cannot determine whether the observed TNF-α pattern represented adaptive inflammation, delayed resolution, or random variation.

Overall, the findings refine rather than fully confirm the original dual-route premise. The study supports a systemic anti-inflammatory contribution of oral *Channa striata*, while the incremental systemic benefit of adding topical *Centella asiatica* was not demonstrated by serum IL-6 measurement. Importantly, this study should be interpreted as an inflammatory biomarker study in a diabetic incisional wound model rather than as direct evidence of improved wound repair. No direct wound-healing endpoints were assessed, such as wound closure rate, wound area reduction, epithelialization, granulation tissue formation, collagen deposition, angiogenesis, or histopathological scoring. Therefore, the present findings indicate changes in selected inflammatory biomarkers, particularly systemic IL-6, but do not demonstrate that the interventions improved structural or functional wound healing. Future combination studies should not rely solely on circulating cytokines but should integrate serial cytokine profiling with local structural and functional endpoints, including wound contraction, re-epithelialization, collagen deposition, angiogenesis, histopathology, and tissue-level inflammatory markers. Such endpoints are needed to determine whether biomarker modulation translates into improved wound repair (20-22).

Study Limitations

This study has several limitations. The sample size was determined using the Federer formula rather than a formal power analysis, which may have limited the ability to detect smaller between-group differences, particularly for wound tissue TNF-α. Outcome measurement was performed with coded samples and blinded assessment, but treatment administration was not blinded because the interventions differed in route and formulation. Extract characterization was limited because specific *Centella asiatica* triterpenoids and *Channa striata* protein components were not quantified. Cytokines were measured at a single endpoint, and serum IL-6 did not differ significantly between G1 and G2, limiting its discriminatory value as a systemic inflammatory marker in this model. No direct structural or functional wound-healing endpoints, such as wound contraction, epithelialization, collagen deposition, angiogenesis, or histopathological scoring, were assessed. Therefore, the observed IL-6 reduction cannot be interpreted as direct evidence of improved tissue repair. Additionally, only male Wistar rats and an incisional wound model were used, limiting generalizability to chronic diabetic ulcers in humans.

CONCLUSION

Oral *Channa striata*, either alone or combined with topical 10% *Centella asiatica* gel, was associated with lower serum IL-6 compared with diabetic vehicle treatment in this Wistar rat diabetic incisional wound model. Wound tissue TNF-α did not differ significantly among groups, and the combination did not show an additive systemic IL-6 reduction beyond oral *Channa striata* monotherapy.

Ethics

Ethics Committee Approval: Procedures were approved by the Bioethics Committee of Sultan Agung Islamic University, Semarang, Indonesia (no: 671/XII/2025/Komisi Bioetik, date: 15.12.2025).

Informed Consent: Patient consent was not applicable because this experimental study involved animals and did not include human participants.

Footnotes

Authorship Contributions

Surgical and Medical Practices: D.I.N., E.K., Concept: D.I.N., E.K., J.W.W., Design: D.I.N., E.K., J.W.W., Data Collection or Processing: D.I.N., Analysis or Interpretation: D.I.N., E.K., J.W.W., Literature Search: D.I.N., E.K., Writing: D.I.N.

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

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

Supplementary Table S1: https://d2v96fxpocvxx.cloudfront.net/ae7a798f-afa1-42f4-9efd-6a2ed974293f/content-images/e1aca7eb-de59-4997-9834-9e23cf063882.pdf

REFERENCES

1. Burgess JL, Wyant WA, Abdo Abujamra B, Kirsner RS, Jozic I. Diabetic wound-healing science. Medicina (Kaunas). 2021; 57: 1072.
2. Dasari N, Jiang A, Skochdopole A, Chung J, Reece EM, Vorstenbosch J, et al. Updates in diabetic wound healing, inflammation, and scarring. Semin Plast Surg. 2021; 35: 153-8.
3. Chen J, Qin S, Liu S, Zhong K, Jing Y, Wu X, et al. Targeting matrix metalloproteases in diabetic wound healing. Front Immunol. 2023; 14.
4. Kuo YS, Chien HF, Lu W. Plectranthus amboinicus and Centella asiatica cream for the treatment of diabetic foot ulcers. Evid Based Complement Alternat Med. 2012; 2012: 418679.
5. Liu YQ, Zhang D, Deng J, Liu Y, Li W, Nie X. Preparation and safety evaluation of Centella asiatica total glycosides nitric oxide gel and its therapeutic effect on diabetic cutaneous ulcers. Evid Based Complement Alternat Med. 2022; 2022: 1419146.
6. Arribas-López E, Zand N, Ojo O, Snowden MJ, Kochhar T. A systematic review of the effect of *Centella asiatica* on wound healing. Int J Environ Res Public Health. 2022; 19: 3266.
7. Witkowska K, Paczkowska-Walendowska M, Garbiec E, Cielecka-Piontek J. Topical application of *Centella asiatica* in wound healing: recent insights into mechanisms and clinical efficacy. Pharmaceutics. 2024; 16: 1252.
8. Idramsyah I, Dahrizal D, Ruran M, Nair HKR. Enhancing Diabetic foot ulcer healing: the role of Channa striata extract nutrition in accelerating inflammatory, proliferative, and maturation phases. IJNP (Indonesian Journal of Nursing Practices). 2024; 8: 99-104.
9. Ramadhanti NA, Sandhika W, Widodo ADW. The effect of snakehead fish (Channa striata) extract on inflammation reaction of skin wound tissue in rattus novergicus wistar strain. Berkala Ilmu Kesehatan Kulit dan Kelamin. 2021; 33: 48.
10. Abd Aziz Syed Hassan S, Rosnelifaizur R, Mazuwin Maya Y, Mutum Samarendra S, Saringat B, Nyi Naing N, et al. Effect of oral Channa striatus in laparotomy wound healing of malnourished rats. International Medical Journal. 2020; 27: 67-70.
11. Zamanifard M, Nasiri M, Yarahmadi F, Zonoori S, Razani O, Salajegheh Z, et al. Healing of diabetic foot ulcer with topical and oral administrations of herbal products: a systematic review and meta-analysis of randomized controlled trials. Int Wound J. 2024; 21: e14760. Erratum in: Int Wound J. 2024; 21: e14875.
12. Rai V, Moellmer R, Agrawal DK. Clinically relevant experimental rodent models of diabetic foot ulcer. Mol Cell Biochem. 2022; 477: 1239-47.
13. Hu Y, Xiong Y, Tao R, Xue H, Chen L, Lin Z, et al. Advances and perspective on animal models and hydrogel biomaterials for diabetic wound healing. Biomater Transl. 2022; 3: 188-200.
14. Clayton SM, Shafikhani SH, Soulika AM. Macrophage and neutrophil dysfunction in diabetic wounds. Adv Wound Care (New Rochelle). 2024; 13: 463-84.
15. Mirastschijski U, Lupše B, Maedler K, Sarma B, Radtke A, Belge G, et al. Matrix metalloproteinase-3 is key effector of TNF-α-induced collagen degradation in skin. Int J Mol Sci. 2019; 20: 5234.
16. Rangwala US, Tashrifwala F, Egbert NN, Asif AA. The potential of topical therapy for diabetic wounds: a narrative review. Cureus. 2023; 15: e36887.
17. Kreiner FF, Kraaijenhof JM, von Herrath M, Hovingh GKK, von Scholten BJ. Interleukin 6 in diabetes, chronic kidney disease, and cardiovascular disease: mechanisms and therapeutic perspectives. Expert Rev Clin Immunol. 2022; 18: 377-89.
18. Furman BL. Streptozotocin-induced diabetic models in mice and rats. Curr Protoc. 2021; 1: e78.
19. Tylutka A, Walas Ł, Zembron-Lacny A. Level of IL-6, TNF, and IL-1β and age-related diseases: a systematic review and meta-analysis. Front Immunol. 2024; 15: 1330386.
20. Mamun AA, Shao C, Geng P, Wang S, Xiao J. Recent advances in molecular mechanisms of skin wound healing and its treatments. Front Immunol. 2024; 15: 1395479.
21. Mahmoud NN, Hamad K, Al Shibitini A, Juma S, Sharifi S, Gould L, et al. Investigating inflammatory markers in wound healing: understanding implications and identifying artifacts. ACS Pharmacol Transl Sci. 2024; 7: 18-27.
22. Gao M, Guo H, Dong X, Wang Z, Yang Z, Shang Q, et al. Regulation of inflammation during wound healing: the function of mesenchymal stem cells and strategies for therapeutic enhancement. Front Pharmacol. 2024; 15: 1345779.

Original Investigation | Özgün Araştırma

Gazi Med J 2026;37(4):475-481

DOI: http://dx.doi.org/10.12996/gmj.2026.4681

Investigation of the Level of Awareness and Knowledge of Physicians Regarding the Cancer Susceptibility of Germline *BRCA1* and *BRCA2* Mutations

Germline *BRCA1* ve *BRCA2* Mutasyonlarının Neden Olduğu Kanser Yatkınlığına Dair Hekimlerin Farkındalık ve Bilgi Düzeylerinin İncelenmesi

Sinan Kolukısaoğlu[1], Berru Atlıhan[1], Mustafa Sırdaş Yılmaz[1], Emre Kağan Dizen[1], Beyza Doğanay[2], Şule Altıner[3]

[1]Ankara University, Faculty of Medicine, Ankara, Türkiye
[2]Department of Biostatistics, Ankara University, Faculty of Medicine, Ankara, Türkiye
[3]Department of Medical Genetics, Ankara University, Faculty of Medicine, Ankara, Türkiye

ABSTRACT

Objective: The rising incidence and mortality of breast cancer highlight the importance of awareness of *BRCA1/2*-related hereditary breast and ovarian cancer. Appropriate genetic understanding and counseling improve outcomes in disease management.

Methods: This study aimed to evaluate the knowledge and awareness levels of physicians at the Ankara University, Faculty of Medicine using a questionnaire administered to 110 physicians.

Results: Among physicians, 47.7% reported adequately informing patients with *BRCA* mutations, whereas most reported difficulties in accessing genetic testing. Significant associations were observed between specialty knowledge and guideline knowledge, and between self-reported knowledge and actual knowledge. Physicians' ability to identify *BRCA*-associated cancers was correlated with both physicians' specialty and education level.

Conclusion: The results indicate a mismatch between perceived and actual knowledge levels and reveal gaps in both awareness and access to genetic services. The study underlines the need for improved education and integration of *BRCA*-related knowledge into clinical practice, emphasizing that medical education and clinical training in cancer genetics are crucial for early diagnosis and effective patient counseling, benefiting both current and future healthcare professionals.

Keywords: *BRCA1* and *BRCA2*, breast cancer, knowledge and awareness, physician

ÖZ

Amaç: Meme kanserinin artan insidansı ve mortalitesi, *BRCA1/2* ile ilişkili kalıtsal meme ve over kanseri farkındalığının önemini vurgulamaktadır. Uygun genetik yaklaşım ve danışmanlık, hastalık yönetimindeki klinik sonuçları iyileştirmektedir.

Yöntemler: Bu çalışma, 110 hekime uygulanan bir anket aracılığıyla Ankara Üniversitesi, Tıp Fakültesi'nde görev yapan hekimlerin bilgi ve farkındalık düzeylerini değerlendirmeyi amaçlamıştır.

Bulgular: Hekimlerin %47,7'si *BRCA* mutasyonu taşıyan hastaları yeterince bilgilendirdiğini bildirirken, büyük bir çoğunluğu genetik testlere erişimde zorluk yaşadıklarını ifade etmiştir. Uzmanlık alanı bilgisi ile kılavuz bilgisi arasında ve öz bildirime dayalı bilgi düzeyi ile gerçek bilgi düzeyi arasında anlamlı ilişkiler gözlemlenmiştir. Hekimlerin BRCA ile ilişkili kanserleri tanımlama yeteneğinin hem uzmanlık alanları hem de eğitim düzeyleri ile korelasyon olduğu bulunmuştur.

Sonuç: Sonuçlar, algılanan ve gerçek bilgi düzeyleri arasında bir uyumsuzluk olduğuna işaret etmekte, ayrıca hem farkındalık hem de genetik hizmetlere erişim konularındaki eksiklikleri ortaya koymaktadır. Bu çalışma, *BRCA* ile ilişkili bilgilerin klinik uygulamaya entegre edilmesine ve daha iyi bir eğitime duyulan ihtiyacın altını çizmektedir. Kanser genetiği alanındaki tıp eğitiminin ve klinik eğitimin, erken teşhis ve etkili hasta danışmanlığı açısından kritik bir öneme sahip olduğu ve hem mevcut hem de gelecekteki sağlık profesyonellerine fayda sağlayacağı vurgulanmaktadır.

Anahtar Sözcükler: *BRCA1* ve *BRCA2*, meme kanseri, bilgi ve farkındalık, hekim

Cite this article as: Kolukısaoğlu S, Atlıhan B, Yılmaz MS, Dizen EK, Doğanay B, Altıner Ş. Investigation of the level of awareness and knowledge of physicians regarding the cancer susceptibility of germline BRCA1 and BRCA2 mutations. Gazi Med J. 2026;37(4):475-481

Address for Correspondence/Yazışma Adresi: Şule Altıner, Department of Medical Genetics, Ankara University, Faculty of Medicine, Ankara, Türkiye
E-mail / E-posta: bicers@ankara.edu.tr
ORCID ID: orcid.org/0000-0001-5789-8630

Received/Geliş Tarihi: 04.04.2026
Accepted/Kabul Tarihi: 23.08.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

INTRODUCTION

Breast carcinoma is the leading cause of cancer mortality in women in Türkiye, mainly due to lack of awareness and delayed diagnosis (1). *BRCA1* and *BRCA2* mutations significantly increase lifetime breast cancer risk (2) and hereditary breast and ovarian cancer (HBOC) syndrome accounts for approximately 10% of breast cancer cases (3). "The *BRCA1* gene, located on chromosome 17, and the *BRCA2* gene, located on chromosome 13, encode proteins that (i) regulate the transcription of other genes, (ii) control the cell cycle and (iii) repair breaks in DNA through homologous recombination (4).

When *BRCA1* and *BRCA2* genes are mutated, they can lose their control function and become responsible for the individual's greater hereditary predisposition to neoplastic development. Breast and ovarian cancers are the most frequently occurring neoplasms associated with pathogenic variants of the *BRCA1* and *BRCA2* genes (5). Inherited disease-causing variants in *BRCA1/2* increase risks of breast, ovarian, fallopian tube, primary peritoneal cancers, and in men, breast and prostate cancers; both sexes face elevated pancreatic cancer risk (6).

Genetic testing involves risk assessment through counseling regarding family cancer history, thereby guiding testing decisions. About 13% of women in the general population develop breast cancer, while 55-72% with disease causing *BRCA1* variants and 45-69% with *BRCA2* variants develop breast cancer by age 70-80 (7-9). Risk varies according to individual factors. Women with *BRCA1/2* variants face higher risk of contralateral breast cancer, increasing to 20-30% at 10 years and 40-50% at 20 years post-diagnosis (8-10). The cumulative risk of all types of skin cancer from age 20 to 80 years is shown to be 14.1% for *BRCA1* carriers and 10.7% for *BRCA2* carriers (11).

BRCA1/2 associated cancer susceptibility is relatively high in Türkiye when compared with United States and European countries (3-4% in Europe, up to 15% in United States) (12), according to a study, 15.1% of Turkish patients with a family history of breast and/or ovarian cancer to have *BRCA1* and *BRCA2* mutations (13).

The education of physicians regarding breast cancer and *BRCA*-related cancer can be limited by several factors, including high workload in outpatient clinics, lack of awareness and initiatives focused on breast cancer and patient outcomes, and the absence of comprehensive local guidelines accessible to physicians from different specialties.

Our study revealed a significant lack of knowledge and awareness among general practitioners and specialists regarding genetic counseling for *BRCA*-related cancers, highlighting the need for more education on genetic counseling and for guidelines on referral to genetic testing in Türkiye.

This study is one of the first studies in Türkiye to include an objective assessment of physicians' knowledge and awareness regarding *BRCA1/2* germline mutations. Identifying gaps in understanding of hereditary breast and ovarian cancer would improve medical education and healthcare. Conducted at a leading academic medical center with diverse departments and access to a broad sample of healthcare professionals and trainees, the study is further strengthened by the institution's longstanding commitment to academic excellence, interdisciplinary research, and public health education, enhancing its relevance and potential impact.

MATERIALS AND METHODS

In this study, physician volunteers completed a structured, validated questionnaire to assess their knowledge and awareness of the cancer susceptibility associated with *BRCA1/2* mutations. Surveys were administered online to volunteers, and the data obtained were analyzed using statistical methods. Ethical approval of this study is done by Ankara University, Faculty of Medicine Student Ethics Committee E-72189195-050.03.04-1402268 (date: 13.06.2024, number: 351065). Informed consent were taken from each participant before collecting data.

Methods for Analysis of Questionnaire

For the analysis of "knowledge evaluation questions" in the survey, the possible answers were assigned points prior to analysis according to their relevance to breast cancer guidelines and the literature, thereby making the assessment of knowledge level more objective. Points can be seen in the following table (Table 1). Significant results for the study's outcomes are presented in tables of percentages.

Inclusion and Exclusion Criteria

A cross-sectional study was conducted amongresidents and specialists in oncology, internal medicine, gynecology and obstetrics, intensive care, medical genetics, oncologic surgery, and general surgery at Ankara University, Faculty of Medicine Hospitals between May and June 2025. Physicians from other specialties are excluded from the study.

Statistical Analysis

Statistical analyses to compare different groups and responses were performed using IBM® SPSS (27th version, IBM Corp., 2020, CA, USA). Chi-square tests and Fisher's exact test were used for group and response comparisons, while independent-samples Kruskal-Wallis, One-Way ANOVA, post-hoc Tukey-Duncan, and correlation tests were used to assess question reliability and guide corrections across several questions and variables. When more than 20% of expected cell counts in a data group were less than five, the Fisher-Freeman-Halton exact test with Monte Carlo simulation was used.

RESULTS

The results revealed a statistically significant association between area of specialization and perceived knowledge levels, indicating that these levels differ among specialties.

Higher mean knowledge levels (mean 3.68) were observed in medical genetics, oncologic surgery, and medical oncology specialists, whereas lower levels were observed in gynecologists (Table 2).

61.4% of participants indicated having good knowledge, when measured their knowledge levels, the 42.2% were at intermediate or lower level about mutation type and cancer matching.

There is a significant difference in the distribution of physicians' experience levels and their stated levels of knowledge regarding *BRCA*-related carcinogenesis (Table 3).

Participants' stated knowledge of breast cancer risk factors differed significantly across responses related to *BRCA1/2* mutation-associated cancers ($p < 0.05$) (Figure 1).

Table 1. The scoring criteria were primarily derived from the NCCN Guidelines, while the Manchester scoring system was used as supplementary background information.

Questions	Answers	Points (question number. points)
Indicate the key diagnostic and prognostic criteria in breast cancer (3,14)	1. Triple breast cancer diagnosis before age of 60	1. +3
	2. Lobular breast cancer seen with diffuse gastric cancer	2. +2
	3. Ancestry of Ashkenazi Jew	3. +2
	4. Two or more than two primary breast cancer in the same person	4. +2
	5. Breast, ovarian, pancreas or/and agressive prostate cancer in 3 or more than 3 relatives including patient	5. +3
Indicate the *BRCA1* and *BRCA2* related cancers (15,16)	1. Breast cancer	1. +1
	2. Ovarian cancer	2. +1
	3. Prostate cancer	3. +1
	4. Pancreatic cancer	4. +1
	5. Melanoma	5. +1
	6. Colorectal cancers	6. 0
	7. Lung cancer	7. 0
	8. Leukemia	8. 0
In which pathological types of breast cancer would you think about hereditary breast-ovarian cancer syndrome?	1. Triple negative	1. +3
	2. HR+	2. +2
	3. Ductal	3.+2
	4. Luminal A	4. +1
	5. Luminal B	5. +1

NCCN: National Comprehensive Cancer Network, HR: Hormone receptor.

Table 2A. Evaluate your level of knowledge regarding the evidence and recommendation levels in *BRCA1/2* genetic testing guidelines.

Specialty/evaluation	1	2	3	4	5	Total
Oncologic surgery	0	0	1	4	4	9
General surgery	2	4	10	10	7	33
General internal medicine	3	3	7	8	0	21
Gynecology and obstetrics	1	3	6	7	7	24
Medical genetics	0	0	0	3	7	10
Medical oncology	0	1	2	4	6	13
Total	**6**	**11**	**26**	**36**	**31**	**110**

Table 2B. Chi-square tests.

	Value	df	Asymptotic significance (2-sided)	Monte Carlo sig. (2-sided)		
				Significance	99% confidence interval	
					Lower bound	Upper bound
Pearson chi-square	78,863[a]	40	0.000	0.047[b]	0.030	0.064
Likelihood ratio	78,952	40	0.000	0.000[b]	0.000	0.005
Fisher-Freeman-Halton exact test	77,486			0.000[b]	0.000	0.005
N of valid cases	110					

[a]: Forty-seven cells (87.0%) have expected count less than 5. The minimum expected count is 0.02, [b]: Based on 1000 sampled tables with starting seed 1314643744.

Table 3. Relations of objectively evaluated and stated knowledge levels.

Objective (vertical, 0-5)/stated (transverse, 1-5)	1	2	3	4	5	Total
0	0	0	0	0	1	1
1	1	0	0	1	2	4
2	0	4	8	6	2	20
3	1	7	7	4	2	21
4	1	2	8	10	7	28
5	0	0	3	14	18	35
Total	**3**	**13**	**26**	**35**	**32**	**109**

Chi-Square Tests

			Asymptotic Significance (2-sided)	Monte Carlo Sig. (2-sided)			Monte Carlo Sig. (1-sided)		
					99% Confidence Interval			99% Confidence Interval	
	Value	df		Significance	Lower Bound	Upper Bound	Significance	Lower Bound	Upper Bound
Pearson Chi-Square	34,807[a]	20	,021	,046[b]	,029	,063			
Likelihood Ratio	30,181	20	,067	,042[b]	,026	,058			
Fisher-Freeman-Halton Exact Test	30,191			,032[b]	,018	,046			
Linear-by-Linear Association	7,993[c]	1	,005	,004[b]	,000	,009	,003[b]	,000	,007
N of Valid Cases	110								

a. 20 cells (66,7%) have expected count less than 5. The minimum expected count is ,02.

b. Based on 1000 sampled tables with starting seed 1502173562.

c. The standardized statistic is 2,827.

Figure 1. A positive association between perceived knowledge and accurate identification of BRCA-related cancers*

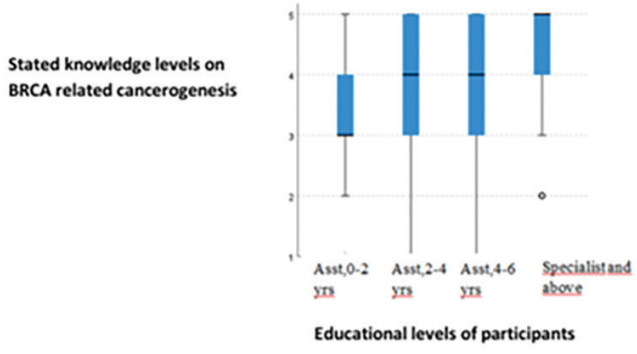

Figure 2. Figure 1 shows the Kruskal-Wallis distribution of knowledge on breast cancer risk factors and evaluated knowledge levels on *BRCA* mutations. Points are given according to Table 1 (0 to 5 points), while self-assessment is done from 1 (very low) to 5 (very high).

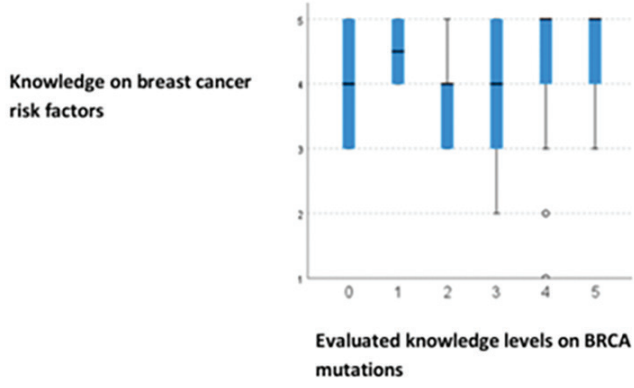

Figure 3. Shows the Kruskal-Wallis distribution of stated knowledge levels on *BRCA*-related carcinogenesis and educational levels of participants. Self-assessment is done from 1 (very low) to 5 (very high) and educational level is from demographic data.

Figure 1 shows the Kruskal-Wallis distribution of knowledge about breast cancer risk factors and evaluated knowledge levels regarding *BRCA* mutations. Points are assigned according to Table 1 (0 to 5 points), and self-assessment is rated on a scale from 1 (very low) to 5 (very high).

There is a significant difference in the distribution of physicians' experience levels and stated levels of knowledge regarding *BRCA*-related carcinogenesis (asymptotic significance = 0.024) (Figure 2).

There is a statistically significant difference in participants' stated knowledge of breast cancer risk factors across their responses related to *BRCA1/2* mutation-associated cancers ($p < 0.05$)

There was no significant relationship between physicians' information-giving attitudes and their knowledge levels regarding breast cancer (Table 4).

Five of six physicians report difficulty accessing genetic testing.

47.7% of physicians stated that they appropriately inform their patients with *BRCA1/2* mutations. 7.2% of physicians stated that they almost never inform their patients about BRCA mutations.

Only 17.3% of respondents believe that breast cancer awareness initiatives are fully sufficient, while 20% of respondents find them clearly insufficient.

DISCUSSION

Our findings demonstrate that, while there is a general understanding of hereditary cancer concepts among participants, significant gaps remain in practical knowledge, application, and awareness, particularly concerning genetic testing guidelines, cancer association, and risk communication.

It was seen that large proportions of the participants reported little or no awareness of genetic testing and services, demonstrating a need to educate physicians about awareness on genetic conditions of their patients (17), but in contrast, in our study, the stated knowledge levels are relatively high when compared with evaluated knowledge levels. These findings indicate that the physicians in our cohort perceive their knowledge to be higher than their actual knowledge.

While most participants rated themselves as having good knowledge, objective assessments showed that many participants lacked adequate understanding, especially regarding mutation-specific cancer associations. Although 32.4% of participants achieved full scores on *BRCA*-related cancer knowledge, a larger proportion of participants remained at intermediate or lower levels. This discordance between perceived and actual knowledge is significant, as it may lead to overconfidence and misinformed clinical decisions**.** Other studies revealed that there is a low level of perceived knowledge indicating a need for a further education for physicians (18) in context of increasing importance of up to date knowledge (19).

Appropriate risk evaluation in genetic counseling is important both to prevent unnecessary genetic testing and to contribute to positive prognostic outcomes (20). In our study, the majority of respondents strongly agreed with the need to inform at-risk relatives of *BRCA1/2* mutation carriers. In this setting, approximately half of physicians stated that they inform their patients with *BRCA1/2* mutations appropriately, and 7.2% of physicians stated that they almost never inform their patients about *BRCA* mutations.

Our study found that specialization significantly influenced knowledge levels. Participants in medical genetics, medical oncology, and oncologic surgery have greater awareness and understanding of *BRCA* testing guidelines and related cancers. This suggests that specialty-specific exposure and training shape familiarity with hereditary cancer concepts. Similarly, knowledge improved with clinical experience but not necessarily with educational attainment, underscoring the value of practical, case-based learning in genetics education. As it becomes more important to support the collaboration of genetic services in the clinic (21). However, in the evaluation of *BRCA1/2*-related cancers, gynecologist participants scored lower than geneticists and other specialists. It can be an indicator for these groups to increase their knowledge on *BRCA* related cancers other than ovarian cancer, which is an important eligibility criterion for hereditary cancer syndromes related with BRCA mutations (22).

Referring patients to medical genetics departments because collaboration between medical genetics and other specialties is an important aspect of clinical genetics practice.

Table 4A. Do you contact with your patients about their genetic conditions? (yes/no).

Objective evaluation points/answers	Yes	No	Total
0	1	1	2
1	4	0	4
2	18	2	20
3	19	1	20
4	27	0	27
5	29	6	35
Total	**98**	**10**	**108**

Table 4B. Do you contact with your patients about their genetic conditions? (yes/no).

	Value	df	Asymptotic significance (2-sided)	Exact sig. (2-sided)
Pearson chi-square	10,148[a]	5	0.071	0.071
Likelihood Ratio	10,848	5	0.054	0.057
Fisher-Freeman-Halton exact test	9,058			0.066
N of valid cases	108			

[a]: Eight cells (66.7%) have expected count less than 5. The minimum expected count is 0.19.

No relationship was observed between physicians' information-giving attitudes and knowledge levels. Even though in the literature, a correlation was shown (23), our study does not indicate for such correlation and our findings show that it become rather a personal choice of physician during practice, who are from different specialties, experience and knowledge levels.

The majority of physicians stated that there should be a referral guideline for medical genetics, indicating a need for such a guideline in Türkiye. Guidelines are important for physicians to guide patient better and make more comprehensive evaluations (20).

Study Limitations

Despite these findings, the study has limitations. Because it is a cross-sectional survey based on self-reported data, it is subject to recall and social desirability bias. Actual clinical behaviors were not directly observed. While we assessed general awareness and perceived competence, more in-depth evaluations of clinical application and guideline adherence would be beneficial for future research. Even points that are given to answers are used to evaluate and compare the knowledge levels appropriately, only literature and evidence-based data can be used to prepare a evaluation scale due to lack of pre-established system

CONCLUSION

This study highlights the need for more standardized, interdisciplinary education for physicians on hereditary cancers, especially those related to *BRCA* mutations.

Targeted educational programs, collaborations between clinicians and greater emphasis on awareness of physicians on genetic topics are critical steps to close the gap between awareness and action in hereditary cancer prevention and management in future. More breast cancer awareness studies and practice-based studies will underscore the implications of knowledge about *BRCA* mutations and breast cancer, leading to applications that yield more positive outcomes for patients. Lifelong learning initiatives regarding the practice of genetics in different fields of clinical medicine, along with increased awareness and knowledge of breast-ovarian cancer syndromes, can enable better management of these syndromes through collaboration between the healthcare service team, the patient, and their relatives.

Guidelines on breast cancer and other cancers with related mutations in Türkiye would help physicians make comprehensive evaluations of patients and improve recognition of breast cancer in a more scientific manner.

Ethics

Ethics Committee Approval: Ethical approval of this study is done by Ankara University, Faculty of Medicine Student Ethics Committee E-72189195-050.03.04-1402268 (date: 13.06.2024, number: 351065).

Informed Consent: Informed consent were taken from each participant before collecting data.

Footnotes

Authorship Contributions

Surgical and Medical Practices: Ş.A., Concept: S.K., B.A., M.S.Y., Ş.A., Design: S.K., B.A., M.S.Y., E.K.D., Ş.A., Data Collection or Processing: S.K., M.S.Y., B.D., Ş.A., Analysis or Interpretation: S.K., M.S.Y., B.D., Ş.A., Literature Search: S.K., M.S.Y., E.K.D., Ş.A., Writing: S.K., M.S.Y., Ş.A.

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

Financial Disclosure: This study is supported by the Scientific and Technological Research Council of Türkiye 2209-A program.

REFERENCES

1. Özmen V, Özmen T, Doğru V. Breast cancer in Turkey; an analysis of 20.000 patients with breast cancer. Eur J Breast Health. 2019; 15: 141-6. Erratum in: Eur J Breast Health. 2019; 15: 276.
2. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet. 2003; 72: 1117-30. Erratum in: Am J Hum Genet. 2003; 73: 709.
3. Sessa C, Balmaña J, Bober SL, Cardoso MJ, Colombo N, Curigliano G, et al.; ESMO Guidelines Committee. Electronic address: clinicalguidelines@esmo.org. Risk reduction and screening of cancer in hereditary breast-ovarian cancer syndromes: ESMO Clinical Practice Guideline. Ann Oncol. 2023; 34: 33-47.
4. Yoshimura A, Imoto I, Iwata H. Functions of breast cancer predisposition genes: implications for clinical management. Int J Mol Sci. 2022; 23: 7481.
5. Ponti G, De Angelis C, Ponti R, Pongetti L, Losi L, Sticchi A, et al. Hereditary breast and ovarian cancer: from genes to molecular targeted therapies. Crit Rev Clin Lab Sci. 2023; 60: 640-50.
6. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Blom MJ, et al.; BRCA1 and BRCA2 Cohort Consortium. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. JAMA. 2017; 317: 2402-16.
7. Owens DK, Davidson KW, Krist AH, Barry MJ, Cabana M, et al.; US Preventive Services Task Force. Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer: US preventive services task force recommendation statement. JAMA. 2019; 322: 652-65. Erratum in: JAMA. 2019; 322: 1830.
8. Finch A, Beiner M, Lubinski J, Lynch HT, Moller P, Rosen B, et al.; Hereditary Ovarian Cancer Clinical Study Group. Salpingo-oophorectomy and the risk of ovarian, fallopian tube, and peritoneal cancers in women with a BRCA1 or BRCA2 Mutation. JAMA. 2006; 296: 185-92.
9. Levine DA, Argenta PA, Yee CJ, Marshall DS, Olvera N, Bogomolniy F, et al. Fallopian tube and primary peritoneal carcinomas associated with BRCA mutations. J Clin Oncol. 2003; 21: 4222-7.
10. Narod SA, Metcalfe K, Finch A, Chan AW, Armel SR, Aeilts A, et al.; Hereditary Breast Cancer Clinical Research Group. The risk of skin cancer in women who carry BRCA1 or BRCA2 mutations. Hered Cancer Clin Pract. 2024; 22: 7.
11. Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol. 2007; 25: 1329-33.
12. Armstrong N, Ryder S, Forbes C, Ross J, Quek RG. A systematic review of the international prevalence of *BRCA* mutation in breast cancer. Clin Epidemiol. 2019; 11: 543-61.

13. Yazici H, Bitisik O, Akisik E, Cabioglu N, Saip P, Muslumanoglu M, et al. BRCA1 and BRCA2 mutations in Turkish breast/ovarian families and young breast cancer patients. Br J Cancer. 2000; 83: 737-42.
14. National Comprehensive Cancer Network. (2025). Genetic/familial high-risk assessment: Breast, ovarian, and pancreatic (version 3.2025). NCCN Clin. Pract. Guidel. Oncol.
15. Sønderstrup IMH, Jensen MR, Ejlertsen B, Eriksen JO, Gerdes AM, Kruse TA, et al. Subtypes in BRCA-mutated breast cancer. Hum Pathol. 2019; 84 :192-201.
16. Vogel VG. 15-epidemiology of breast cancer. Breast. 2018; 207-18. e4.
17. Diamonstein C, Stevens B, Shahrukh Hashmi S, Refuerzo J, Sullivan C, Hoskovec J. Physicians' awareness and utilization of genetic services in Texas. J Genet Couns. 2018; 27: 968-77.
18. Vande Perre P, Toledano D, Corsini C, Escriba E, Laporte M, Bertet H, et al. Role of the general practitioner in the care of BRCA1 and BRCA2 mutation carriers: general practitioner and patient perspectives. Mol Genet Genomic Med. 2018; 6: 957-65.
19. Patch C, Middleton A. Genetic counselling in the era of genomic medicine. Br Med Bull. 2018; 126: 27-36.
20. Nelson HD, Pappas M, Zakher B, Mitchell JP, Okinaka-Hu L, Fu R. Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer in women: a systematic review to update the U.S. preventive services task force recommendation. Ann Intern Med. 2014; 160: 255-66.
21. Slomp C, Morris E; GenCOUNSEL Study; Price M, Elliott AM, Austin J. The stepwise process of integrating a genetic counsellor into primary care. Eur J Hum Genet. 2022; 30: 772-81. Erratum in: Eur J Hum Genet. 2022; 30: 865.
22. Kemp Z, Turnbull A, Yost S, Seal S, Mahamdallie S, Poyastro-Pearson E, et al. Evaluation of cancer-based criteria for use in mainstream BRCA1 and BRCA2 genetic testing in patients with breast cancer. JAMA Netw Open. 2019; 2: e194428.
23. Douma KF, Smets EM, Allain DC. Non-genetic health professionals' attitude towards, knowledge of and skills in discussing and ordering genetic testing for hereditary cancer. Fam Cancer. 2016; 15: 341-50.

Original Investigation | Özgün Araştırma

Gazi Med J 2026;37(4):482-489

DOI: http://dx.doi.org/10.12996/gmj.2026.4713

Genetic Diagnosis and Medical Genetics in Turkish Court of Cassation Decisions: A Cross-sectional Decision Analysis

Türk Yargıtay Kararlarında Genetik Tanı ve Tıbbi Genetik: Kesitsel Bir Karar Analizi

Dilsu Dicle Erkan, Elifcan Taşdelen

Department of Medical Genetics, University of Health Sciences Türkiye, Ankara Etlik City Hospital, Ankara, Türkiye

ABSTRACT

Objective: Advances in genetic and genomic technologies have expanded the clinical role of medical genetics and created complex legal, ethical, and social implications. Despite the growing integration of genetic information into clinical practice and legal proceedings, the representation of genetic and medical genetic concepts in Turkish Court of Cassation decisions has not been systematically examined.

Methods: A retrospective, descriptive, cross-sectional decision analysis was conducted using the publicly accessible Yargıtay Karar Arama database. Two keywords, "genetic diagnosis" and "medical genetics," were searched in Turkish. After removal of duplicates and assessment of eligibility, 16 decisions were included in the final analysis. Decisions were classified into a main analysis group (n = 12) and a secondary analysis group (n = 4). Data were extracted on legal context, genetic themes, role of genetic evidence, expert assessments, and judicial outcomes.

Results: The included decisions spanned 2009 to 2024 and covered both civil and criminal chambers. The most frequent legal context was DNA-based parentage and family law disputes (37.5%), followed by prenatal diagnosis or medical genetics-related malpractice and laboratory responsibility (31.3%), medical genetics service reimbursement disputes (18.8%), and genetic disease-related procedural rights and personality rights claims (12.5%). Expert evaluation was a recurring determinant of judicial reasoning across all categories.

ÖZ

Amaç: Genetik ve genomik teknolojilerdeki gelişmeler, tıbbi genetiğin klinik rolünü genişletmiş ve karmaşık hukuki, etik ve sosyal sonuçlar doğurmuştur. Genetik bilginin klinik uygulamaya ve hukuki süreçlere giderek artan entegrasyonuna karşın, genetik ve tıbbi genetik kavramlarının Türkiye Yargıtay kararlarındaki yansımaları sistematik olarak incelenmemiştir.

Yöntemler: Kamuya açık Yargıtay Karar Arama veri tabanı kullanılarak geriye dönük, tanımlayıcı, kesitsel bir karar analizi gerçekleştirilmiştir. "Genetik tanı" ve "tıbbi genetik" olmak üzere iki anahtar kelime Türkçe olarak taranmıştır. Yinelenen kararların çıkarılması ve uygunluk değerlendirmesinin ardından 16 karar nihai analize dahil edilmiştir. Kararlar, ana analiz grubu (n = 12) ve ikincil analiz grubu (n = 4) olarak sınıflandırılmıştır. Her karardan hukuki bağlam, genetik temalar, genetik kanıtın rolü, uzman değerlendirmeleri ve yargısal sonuçlara ilişkin veriler elde edilmiştir.

Bulgular: Dahil edilen kararlar 2009-2024 yılları arasında yayılmakta olup hem hukuk hem de ceza dairelerini kapsamaktadır. En sık karşılaşılan hukuki bağlam, DNA temelli soybağı ve aile hukuku uyuşmazlıkları (%37,5) olmuştur. Bunu sırasıyla prenatal tanı veya tıbbi genetik kaynaklı malpraktis ve laboratuvar sorumluluğu davaları (%31,3), tıbbi genetik hizmeti geri ödeme uyuşmazlıkları (%18,8) ve genetik hastalığa bağlı usul hakları ile kişilik hakkı talepleri (%12,5) izlemiştir. Uzman değerlendirmesi, tüm kategorilerde yargısal muhakemenin tekrarlayan bir belirleyicisi olarak öne çıkmıştır.

Cite this article as: Erkan DD, Taşdelen E. Genetic diagnosis and medical genetics in Turkish Court of Cassation decisions: a cross-sectional decision analysis. Gazi Med J. 2026;37(4):482-489

Address for Correspondence/Yazışma Adresi: Dilsu Dicle Erkan, Department of Medical Genetics, University of Health Sciences Türkiye, Ankara Etlik City Hospital, Ankara, Türkiye
E-mail / E-posta: dilsudicle@gmail.com
ORCID ID: orcid.org/0000-0002-2400-8344

Received/Geliş Tarihi: 14.05.2026
Accepted/Kabul Tarihi: 15.06.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

Conclusion: Decisions of the Turkish Court of Cassation reveal that genetics has become a multidimensional legal issue, extending beyond parentage testing into prenatal diagnosis, laboratory responsibilities, informed consent, reimbursement, and procedural rights. Emerging genomics-related issues such as variant reinterpretation, duties to warn relatives, and genetic discrimination have not yet achieved visibility before the high courts in Türkiye.

Keywords: Medical genetics, genetic diagnosis, Court of Cassation, judicial decisions, medical law, prenatal diagnosis, DNA testing

ÖZ

Sonuç: Yargıtay kararları, genetiğin soybağı tespitinin ötesine geçerek prenatal tanı, laboratuvar sorumluluğu, aydınlatılmış onam, geri ödeme ve usul hakları alanlarına uzanan çok boyutlu bir hukuki mesele haline geldiğini ortaya koymaktadır. Varyantların yeniden yorumlanması, risk altındaki akrabaları uyarma yükümlülüğü ve genetik ayrımcılık gibi genetik ilişkili hukuki sorunlar ise henüz Türkiye'de yüksek mahkemelerin gündemine girmemiştir.

Anahtar Sözcükler: Tıbbi genetik, genetik tanı, Yargıtay, yargı kararları, tıp hukuku, prenatal tanı, DNA analizi

INTRODUCTION

Advances in genetic testing have substantially expanded the role of medical genetics in contemporary medicine. Genetic and genomic technologies are now used not only for the diagnosis of rare inherited disorders, but also in prenatal diagnosis, carrier screening, preimplantation genetic testing, human leukocyte antigen (HLA) matching, pharmacogenetics, cancer susceptibility assessment, and individualized risk prediction (1). As these applications increasingly influence diagnosis, reproductive decision-making, treatment planning, and long-term care, medical genetics has become an integral component of modern health care.

The growing clinical use of genetic information has also created important legal, ethical, and social implications (2). Genetic test results may have direct consequences not only for the individual being tested, but also for family members and future offspring. In legal practice, genetic evidence may serve as decisive proof in parentage and paternity disputes, while prenatal genetic testing may become central to malpractice claims involving missed or delayed diagnosis, inadequate counseling, laboratory error, or failure to obtain informed consent (3-5). Similarly, disputes may arise regarding the reimbursement of genetic tests, the medical necessity of specialized genetic services, and the interpretation of expert reports in cases involving genetic diseases (2). Therefore, genetic information increasingly functions as both a clinical tool and a legal determinant.

High court decisions provide a valuable source for understanding how these complex issues are interpreted within a legal system. In Türkiye, decisions of the Court of Cassation are particularly important because they reflect the judicial approach to causation, standard of care, expert evaluation, informed consent, evidentiary sufficiency, and procedural requirements. Previous decision-based analyses in medical law have demonstrated that court judgments can reveal recurring patterns in malpractice allegations, expert report conflicts, and judicial reasoning (6). Despite the expanding role of medical genetics in clinical practice, how genetic and medical genetic concepts are represented in decisions of the Turkish Court of Cassation has not been systematically examined.

Although genetic information is increasingly used in both clinical practice and legal decision-making, its representation in Turkish Court of Cassation decisions remains unclear. Therefore, this study aimed to systematically examine decisions of the Court of Cassation and to classify their legal contexts, clinical and genetic themes, role of genetic evidence, expert assessments, and judicial outcomes.

MATERIALS AND METHODS

Study Design

This study was designed as a retrospective, descriptive, cross-sectional decision analysis of publicly available decisions of the Turkish Court of Cassation concerning genetic diagnosis and medical genetics. The study aimed to evaluate the legal contexts in which genetic and medical-genetic concepts appear, the role of genetic evidence or genetic services in judicial reasoning, and the outcomes of those decisions.

Data Source and Search Strategy

The data source was the publicly accessible *"Yargıtay Karar Arama"* database (https://karararama.yargitay.gov.tr/#), which contains anonymized decisions of the Turkish Court of Cassation. The database was searched using two predefined keywords: genetic diagnosis (originally searched in Turkish as *"genetik tanı"*) and medical genetics (originally searched in Turkish as *"tıbbi genetik"*).

The keyword "genetic diagnosis" yielded 13 decisions, while "medical genetics" yielded 12 decisions. Thus, 25 records were initially identified. After comparing the search results, 4 duplicate records were removed. The remaining 21 unique decisions were reviewed at the full-text level for eligibility.

Eligibility Criteria

Decisions were included if a genetic or medical genetic element had a meaningful role in the factual background, evidentiary assessment, expert evaluation, legal reasoning, or judicial outcome of the case. Specifically, decisions were considered eligible if they involved at least one of the following:

Genetic testing, DNA analysis, molecular genetic analysis, prenatal genetic diagnosis, preimplantation genetic diagnosis, HLA tissue typing, genetic disease, medical genetics expert assessment, medical genetics laboratory evaluation, or disputes concerning the necessity, reimbursement, or legal consequences of genetic services.

Decisions were excluded when the terms "genetic diagnosis" or "medical genetics" appeared only incidentally (for example, as part of a department or institution name, a device location, an academic position, a commercial contract, or a technical service relationship)

and there was no substantive discussion of genetic testing, genetic disease, medical genetic evaluation, or genetic health service. Duplicate decisions and those that did not contain sufficient relevant content for analysis were also excluded.

After eligibility assessment, 5 decisions were excluded and 16 were included in the final analysis (Figure 1).

The main analysis group included decisions in which genetic testing, genetic disease, medical genetic evaluation, genetic laboratory responsibility, or genetic health service was directly central to the dispute or judicial reasoning. This group included 12 decisions.

The secondary analysis group included decisions in which a genetic test result or genetic disease appeared as background information or supporting evidence, while the principal legal dispute concerned another issue such as divorce, compensation, waiver of a paternity claim, or procedural rights in criminal proceedings. This group included 4 decisions.

Data Extraction

For each included decision, the following variables were extracted into a standardized data collection form:

Court of Cassation chamber, case number, decision number, decision year, type of court, civil or criminal jurisdiction, type of legal dispute, genetic or medical genetic concept involved, clinical or genetic condition, role of the genetic element in the decision, presence of expert report, presence of medical genetics expert or medical genetics department assessment, presence of Forensic Medicine Institute report, presence of conflicting expert reports, first-instance court outcome, regional appellate court outcome if available, Court of Cassation outcome, and main reasoning of the Court of Cassation.

The genetic or medical-genetic concepts were categorized as DNA/parentage testing, molecular genetic analysis, prenatal diagnosis, preimplantation genetic diagnosis, HLA tissue typing, genetic disease, medical-genetics expert evaluation, medical-genetics laboratory responsibility, or medical-genetics service/reimbursement dispute.

The role of the genetic element was categorized as primary evidence, supporting evidence, expert assessment, malpractice or negligence evaluation, informed consent issue, causation assessment, health service necessity/reimbursement issue, procedural rights issue, or background information.

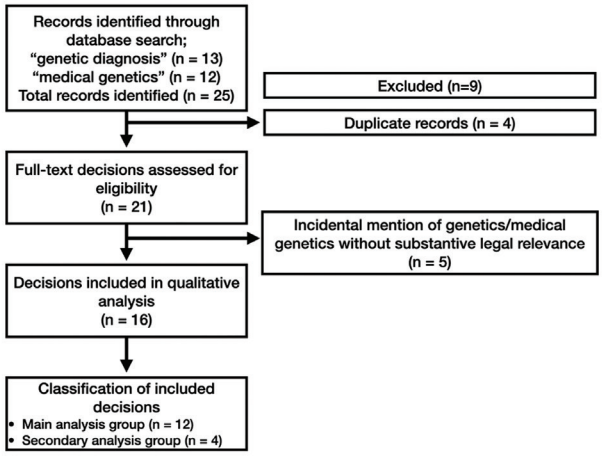

Figure 1. Schematic algorithm for inclusion & exclusion of decisions.

Outcome Measures

The primary outcome was the distribution of legal contexts in which genetic and medical genetic concepts appeared in Court of Cassation decisions. Secondary outcomes included the type of genetic/genomic issue addressed, the role of genetic evidence or medical genetic assessment in judicial reasoning, the presence and nature of expert reports, the existence of conflicting expert opinions, and the final Court of Cassation outcome.

Statistical Analysis

Data were analyzed descriptively. Categorical variables were presented as numbers and percentages. Because of the limited sample size and the descriptive nature of the study, no comparative statistical analysis was performed.

Ethical Considerations

This study did not involve human participants, patient records, direct patient contact, or identifiable personal data. All decisions were obtained from the publicly accessible Yargıtay Karar Arama database and had already been anonymized. Therefore, ethics committee approval and informed consent were not required.

RESULTS

Decision Selection

The initial search identified 25 records from the Turkish Court of Cassation database; 16 decisions were included in the analysis. Of these, 12 decisions were classified in the main analysis group, and 4 in the secondary analysis group.

General Characteristics of Included Decisions

The included decisions were issued between 2009 and 2024 and involved both the civil and criminal chambers of the Court of Cassation. Most decisions were from civil chambers, whereas two were from criminal chambers. The main analysis group included decisions in which genetic testing, medical genetic evaluation, genetic laboratory responsibility, or genetic health services were central to the dispute. The secondary analysis group included decisions in which genetic testing or genetic disease served as background or supporting evidence, while the principal legal dispute involved another issue, such as divorce, compensation, waiver of a paternity claim, or procedural rights in criminal proceedings (Table 1).

The decisions covered a broad range of legal and clinical contexts. The most frequent category was DNA-based parentage and family law disputes, accounting for 6 of the 16 decisions (37.5%). This was followed by prenatal diagnosis or medical genetics-related malpractice and laboratory responsibility, accounting for 5 decisions (31.3%). Disputes concerning reimbursement or financing of medical genetics services were identified in 3 decisions (18.8%). Genetic disease and criminal procedural rights, as well as DNA-related personality rights or compensation disputes, each accounted for 1 decision (6.3%) (Figure 2).

Legal Contexts and Genetic Themes

DNA or parentage testing was the most common genetic theme identified among the included decisions. These cases involved paternity actions, denials of parentage, divorce proceedings based on DNA evidence, waiver-related disputes in paternity litigation, and compensation claims following the discovery of biological non-paternity. In these decisions, DNA testing generally served as a central or highly influential evidentiary element in determining biological relationships or the legal consequences arising from such determinations (Table 2).

Prenatal diagnosis-related disputes constituted the second major group. These decisions involved Down syndrome, prenatal screening, chorionic villus sampling, maternal cell contamination,

Table 1. Included Court of Cassation decisions and classification.

No.	Chamber	Case/decision no.	Year	Analysis group	Legal context	Genetic/clinical theme	Role of genetic element
1	11th civil chamber	2023/5901 E., 2024/9400 K.	2024	Main analysis	Prenatal diagnosis-related malpractice/laboratory responsibility	Down syndrome, CVS, maternal cell contamination	Assessment of laboratory error and medical genetics responsibility
2	11th civil chamber	2023/1669 E., 2024/4009 K.	2024	Main analysis	Prenatal diagnosis and informed consent	Down syndrome, prenatal screening	Evaluation of counseling, informed consent
3	2nd civil chamber	2021/5966 E., 2021/9684 K.	2021	Main analysis	Paternity action	DNA testing	Primary evidence; adequacy of DNA report and requirement for expert committee evaluation
4	2nd civil chamber	2023/1987 E., 2023/2378 K.	2023	Main analysis	Denial of parentage	Molecular genetic analysis	Primary evidence for absence of parentage relationship
5	18th civil chamber	2015/11880 E., 2016/7576 K.	2016	Main analysis	Denial of parentage	Private genetic diagnosis center report	Private DNA report and need for official verification
6	2nd civil chamber	2022/1067 E., 2022/2242 K.	2022	Main analysis	Denial of parentage and limitation period	Genetic-biological origin, out-of-court DNA testing	Determination of the date of knowledge for filing period
7	3rd civil chamber	2023/1162 E., 2023/3046 K.	2023	Main analysis	Prenatal genetic testing-related laboratory responsibility	Thalassemia, Sanger sequencing, mutation analysis	Assessment of laboratory responsibility, causation, and expert reports
8	11th civil chamber	2020/342 E., 2021/6946 K.	2021	Main analysis	Prenatal screening, informed consent, and malpractice	Down syndrome, amniocentesis, prenatal screening tests	Evaluation of medical records, informed consent, and prenatal counseling
9	5th criminal chamber	2012/6668 E., 2013/3008 K.	2013	Main analysis	Criminal proceedings related to alleged medical negligence	Prenatal care, medical genetics expert opinion	Conflicting expert reports and need for further specialist evaluation
10	3rd civil chamber	2020/3411 E., 2020/7269 K.	2020	Main analysis	Reimbursement/deduction dispute for medical genetics services	Medical genetics tests	Assessment of medical necessity of genetic investigations
11	13th civil chamber	2012/10869 E., 2013/4645 K.	2013	Main analysis	Medical genetics service purchase and reimbursement dispute	Medical genetics services	Financial/legal evaluation of medical genetics service provision
12	10th civil chamber	2009/5150 E., 2009/16257 K.	2009	Main analysis	Coverage of health expenses	PGD, HLA tissue typing, IVF, stem cell donor sibling	Evaluation of PGD/HLA-related expenses as part of necessary treatment
13	4th civil chamber	2015/12078 E., 2017/3155 K.	2017	Secondary analysis	Personality rights and compensation following parentage dispute	DNA testing	Background evidence leading to compensation/personality rights dispute
14	8th civil chamber	2020/946 E., 2021/5413 K.	2021	Secondary analysis	Invalidity of waiver in paternity action	DNA testing, genetic diagnosis center report	Background evidence in waiver, fraud, and procedural validity dispute
15	2nd civil chamber	2023/49 E., 2023/2994 K.	2023	Secondary analysis	Divorce based on adultery and related compensation	Private DNA/genetic diagnosis report	Background evidence supporting adultery and parentage-related dispute
16	12th criminal chamber	2017/1050 E., 2018/10198 K.	2018	Secondary analysis	Criminal procedural rights	Huntington disease, molecular genetic analysis	Genetic disease as background for mandatory defense counsel and fair trial concerns

CVS: Chorionic villus sampling, HLA: Human leukocyte antigen, IVF: In vitro fertilization, PGD: Preimplantation genetic diagnosis.

amniocentesis, thalassemia, Sanger sequencing, and mutation analysis. In these cases, genetic or medical-genetic issues were mainly discussed in relation to alleged missed diagnoses, inadequate counseling, informed consent, laboratory error, causation, and medical negligence (Tables 1 and 2).

Three decisions involved reimbursement or financing of medical genetics services. These included disputes regarding the necessity and financial coverage of genetic investigations, medical genetics service provision, and preimplantation genetic diagnosis with HLA tissue typing for the purpose of identifying a potential stem cell donor sibling. These decisions demonstrated that genetic services were relevant not only to malpractice and family law disputes but also to health financing and reimbursement-related litigation (Table 3).

Role of Genetic Evidence and Medical Genetic Assessment

The role of the genetic element varied according to the legal context. In 5 decisions, genetic or DNA testing served as the primary evidence, particularly in parentage or biological relationship disputes. In 3 decisions, genetic testing or prenatal diagnosis was used to assess malpractice, negligence, laboratory responsibility, or causation. In 2 decisions, the genetic issue was closely related to informed consent and counseling, especially regarding prenatal screening, amniocentesis, and the limitations or implications of genetic testing (Table 3).

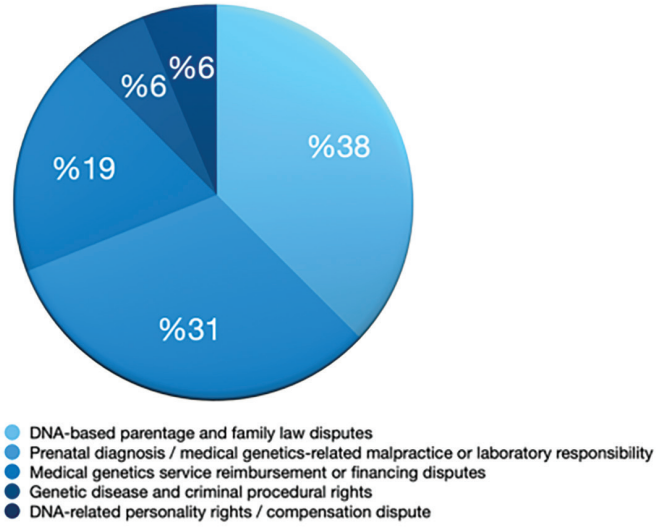

Figure 2. Pie chart showing legal contexts of individual decisions.

In 3 decisions, the genetic element was evaluated with respect to the necessity of health services or reimbursement. These cases focused on whether genetic investigations or interventions related to genetic diagnosis should be considered medically necessary or financially covered. In 1 decision, a genetic disease was relevant to procedural rights in criminal proceedings, specifically regarding the need for mandatory defense counsel. In 2 decisions, genetic evidence served as background or supporting evidence for a related legal claim rather than forming the primary dispute itself (Table 3).

Expert Reports and Judicial Reasoning

Expert evaluation was a recurring theme across the included decisions. In parentage-related cases, the adequacy and evidentiary value of DNA reports were emphasized, including the need for reliable expert assessment when determining biological relationships. In cases related to prenatal diagnosis, medical genetics was commonly associated with expert interpretation of test results, laboratory responsibility, informed consent, and causation. Several decisions highlighted the importance of specialist evaluation, particularly in cases where the dispute involved complex prenatal testing, genetic laboratory procedures, or conflicting expert opinions.

In medical-genetics-related malpractice and laboratory-responsibility cases, expert reports were especially important in distinguishing among an unavoidable limitation of testing, a complication, and a preventable error. The decisions involving prenatal diagnosis and thalassemia testing illustrated how courts may rely on expert assessments to evaluate whether the appropriate genetic test was performed, whether the result was interpreted correctly, whether the patient was adequately informed, and whether an alleged error had a causal relationship with the claimed harm (Tables 1 and 3).

Summary of Findings

Overall, the included cases showed that genetics and medical genetics featured in decisions of the Court of Cassation across multiple legal domains. Although DNA-based parentage disputes represented the largest category, a substantial proportion of decisions involved

Table 2. Genetic and clinical themes identified in the decisions.

Genetic/clinical theme	No. of decisions	Typical legal relevance
DNA or parentage testing	6	Parentage, paternity, divorce, waiver, and compensation-related disputes
Down syndrome and prenatal screening	3	Prenatal counseling, informed consent, and alleged missed diagnosis
Medical genetics service provision/reimbursement	2	Necessity and financial coverage of genetic investigations
Thalassemia and prenatal molecular testing	1	Laboratory responsibility and interpretation of mutation analysis
PGD and HLA tissue typing	1	Coverage of IVF/PGD expenses for potential stem cell donor sibling
Huntington disease and molecular genetic analysis	1	Procedural rights and need for mandatory defense counsel
Prenatal care with medical genetics expert involvement	1	Conflicting expert opinions in alleged medical negligence
DNA-related personality rights dispute	1	Compensation after discovery of biological non-paternity
Total	16	—

HLA: Human leukocyte antigen, IVF: In vitro fertilization, PGD: Preimplantation genetic diagnosis.

Table 3. Role of the genetic element in judicial reasoning.

Role of genetic element	No. of decisions	Description
Primary evidence (3–4–5–6–15)	5	Genetic/DNA test result was central to proving or disproving parentage or biological relationship
Malpractice, negligence, or laboratory responsibility assessment (1–7–9)	3	Genetic testing or prenatal diagnosis was used to evaluate alleged medical error, laboratory error, or causation
Informed consent/counseling issue (2–8)	2	Genetic or prenatal testing raised questions regarding patient counseling, test limitations, and informed consent
Health service necessity or reimbursement (10–11–12)	3	Genetic services were evaluated in relation to medical necessity, payment, or public coverage
Procedural rights/legal capacity-related issue (16)	1	Genetic disease was relevant to the defendant's procedural protection and defense rights
Background or supporting evidence (13–14)	2	Genetic evidence was not the main dispute but formed the factual basis of a related legal claim
Total	16	—

broader medical genetics issues, including prenatal diagnosis, genetic laboratory responsibility, informed consent, reimbursement of genetic services, and genetic disease-related procedural rights. These findings suggest that genetic information in high-court decisions functions not only as biological evidence, but also as a basis for evaluating medical responsibility, counseling, health service coverage, and procedural safeguards.

DISCUSSION

The principal finding of this study was that genetic and medical-genetic concepts in decisions of the Turkish Court of Cassation were not confined to DNA-based parentage disputes but also appeared across a broader range of legal contexts, including prenatal diagnosis, laboratory responsibility, informed consent, reimbursement of genetic services, and procedural rights related to genetic disease. Although DNA-based parentage and family law disputes constituted the largest category, nearly one-third of the included decisions involved prenatal diagnosis or medical genetics-related malpractice, demonstrating that genetics has already entered Turkish high court jurisprudence as a multidimensional legal issue rather than merely a tool for establishing biological relationships. This pattern is consistent with the broader international transition from conventional genetics to genomics, in which expanding genetic technologies increasingly generate legal questions related not only to biological identity, but also to clinical responsibility, counseling, interpretation, and access to services (2).

DNA-based parentage disputes constituted the most frequent legal context in the present sample. This finding is unsurprising, because DNA analysis has long been one of the earliest and most established forensic applications of genetic testing in Türkiye, with Turkish data showing its high discriminative value in disputed paternity cases (7). However, the decisions included in this study also show that the Court of Cassation does not treat DNA results as automatically dispositive; rather, it gives substantial weight to the procedural reliability of the report, including whether the analysis was performed by an appropriate expert committee and whether private reports were supported by official evaluation (Turkish Civil Code Arts. 285-309). Thus, the Turkish jurisprudential approach appears to combine recognition of the scientific strength of DNA evidence with continued emphasis on procedural safeguards in matters affecting civil status and family relationships (8).

The most clinically significant group of decisions involved prenatal diagnosis, including cases related to Down syndrome, procedures such as chorionic villus sampling and amniocentesis, issues such as maternal cell contamination, and prenatal molecular testing for thalassemia. These decisions closely parallel the international literature on wrongful birth claims, in which the central allegation is that inadequate testing, incorrect interpretation, or insufficient counseling deprived the parents of the opportunity to make an informed reproductive decision (5,9). The Turkish cases identified in the present study reflected two recurring dimensions of liability: first, whether appropriate prenatal genetic counseling and informed consent were provided, and second, whether the genetic laboratory process was conducted and interpreted appropriately. This is important because prenatal genetic testing is not simply a technical diagnostic act; it is a process in which the clinical value of the test depends on appropriate patient selection, explanation of residual risk, accurate laboratory practice, and documentation of counseling (2-4). Therefore, the Turkish Court of Cassation decisions in this field are broadly aligned with international medico-legal concerns surrounding prenatal diagnosis and reproductive autonomy (5,9).

Another important observation was that the legal function of genetic information differed markedly across contexts. In parentage disputes, genetic testing generally served as primary evidence; in prenatal diagnosis-related cases, it was used to assess negligence, causation, laboratory responsibility, or informed consent; in reimbursement disputes, it was evaluated with respect to medical necessity and financial coverage; and in one criminal case, a genetic disease was relevant to procedural protection and to the right to counsel. This diversity shows that genetic information should not be conceptualized as a single category of proof, because its legal significance depends on the factual and normative structure of each dispute (2,10). Genetics in judicial reasoning may function as biological evidence, a marker of professional responsibility, a determinant of access to care, or a trigger for procedural safeguards.

The decisions also highlighted the central role of expert assessment in disputes involving genetics and medical genetics. Several judgments turned not merely on the existence of a genetic test

result, but on the adequacy of the expert report, the specialty of the evaluating physicians, the resolution of conflicting reports, and the completeness of the underlying medical documentation. This is particularly relevant in genetics, where the legal assessment may depend on complex issues such as sample identity, maternal contamination, test limitations, analytical performance, and the interpretation of uncertain or evolving findings (1,2). The need for qualified specialist input is likely to become even more important as clinical genetics continues to move toward broader genomic testing and reinterpretation of results over time. From a practical standpoint, these decisions reinforce the importance of careful documentation, appropriate referral, and transparent laboratory reporting in reducing legal vulnerability in genetic care.

The reimbursement-related decisions constituted another noteworthy finding of the study. In addition to disputes about payment for medical genetics services, one decision concerned the coverage of in vitro fertilization combined with preimplantation genetic diagnosis and HLA tissue typing to facilitate the birth of a potential stem-cell donor sibling. Internationally, litigation concerning preimplantation genetic diagnosis has most often focused on misdiagnosis, procedural error, and failure to inform patients of inherent error rates, whereas the Turkish decision identified in the present sample illustrates that genetic technologies may also reach the courts through questions of medical necessity and access to publicly funded care (4,11). This suggests that the legal footprint of genetics extends beyond liability into broader issues of health policy, allocation of resources, and the boundaries of reimbursable treatment.

Notably, several topics that are prominent in the international genetics-law literature did not appear in the present sample. These include duties to warn at-risk relatives, reinterpretation or reanalysis of previously reported variants, and genetic discrimination in employment or insurance (1,6,12). For example, the English case of ABC v St George's Healthcare NHS Trust has prompted detailed discussion of whether healthcare professionals may owe duties of care to relatives at genetic risk, while in the United States cases such as Pate v. Threlkel established important early principles regarding familial disclosure of hereditary risk (6,13,14). Similarly, contemporary literature has emphasized emerging legal uncertainty around variant reinterpretation and ongoing disputes regarding the scope of protection against genetic discrimination (1,12). Their absence from the present Turkish Court of Cassation sample should not be interpreted as evidence that such issues are irrelevant in Türkiye; rather, current high-court visibility appears to remain concentrated in more established domains such as parentage, prenatal diagnosis, laboratory responsibility, and reimbursement. Given that genetic data are uniquely identifying, enduring, and informative about both the individual and biological relatives, privacy and familial implications are likely to become increasingly important in future Turkish legal disputes.

The temporal distribution of the decisions also deserves comment. Although the included judgments spanned 2009–2024, nearly two-thirds were issued between 2019 and 2024, suggesting that disputes involving genetic diagnosis and medical genetics have become more visible in recent years. This pattern may reflect the increasing integration of genetic testing into clinical practice, broader availability of prenatal and reproductive genetic technologies, and the expanding use of DNA-based evidence in family law disputes. However, because the present study relied on publicly available decisions retrieved through selected keywords, this trend should be interpreted as increased visibility within the available jurisprudence rather than as definitive evidence of a rise in litigation frequency.

The present findings have several practical implications. For medical geneticists, these decisions underscore the importance of comprehensive prenatal counseling, explicit discussion of test limitations and residual risk, and complete documentation of informed consent. For genetic laboratories, these cases demonstrate the legal importance of maintaining sample integrity, controlling contamination, selecting appropriate methods, and ensuring clear reporting practices. For courts, the findings support the need for appropriately qualified expert panels when adjudicating disputes that depend on specialized genetic knowledge. For policy-makers, the results suggest that clearer frameworks may eventually be needed not only for reimbursement and access to genetic services, but also for emerging areas such as variant reinterpretation, familial disclosure, and genetic discrimination.

Study Limitations

This study has several limitations: First, the search strategy was based on the keywords "genetic diagnosis" and "medical genetics"; therefore, decisions on related issues that used different terminology may not have been captured. Second, only publicly available Court of Cassation decisions were analyzed, and the study therefore reflects the contents of the accessible database rather than the entire genetics-related litigation in Türkiye. Third, the number of included decisions was limited, and the cases were heterogeneous, precluding comparative statistical analyses. Finally, the temporal analysis was based on decision year rather than filing year, which limits the ability to draw conclusions about changes in litigation incidence over time.

CONCLUSION

Turkish Court of Cassation decisions reveal that genetics has become a multidimensional legal issue, extending beyond parentage testing to include prenatal diagnosis, laboratory responsibility, informed consent, reimbursement, and procedural rights. At present, Turkish high court jurisprudence appears to be concentrated in the established areas of genetic practice, whereas newer genomics-related issues, such as variant reinterpretation, duties to warn relatives, and genetic discrimination have not yet been observed in the sample identified in this study. As genomic medicine continues to expand, both clinical practice and legal doctrine will likely need to adapt to increasingly complex questions concerning genetic responsibility, access, privacy, and family-linked risk.

Ethics

Ethics Committee Approval: Ethical approval was not required for this study, as it involved no human participants, patient data, or identifiable personal information.

Informed Consent: No informed consent was required for this study, as publicly available anonymized decisions were analyzed.

Footnotes

Authorship Contributions

Concept: D.D.E., Design: D.D.E., Data Collection or Processing: D.D.E., Analysis or Interpretation: D.D.E., Literature Search: D.D.E., E.T., Writing: D.D.E., E.T.

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

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

References

1. Sentell ZT, Zawati MH. Ethical, legal, and policy dimensions and contentions for reanalysis and reinterpretation of clinical genetic testing results. Front Genet. 2025; 16: 1685854.
2. Marchant G, Barnes M, Evans JP, LeRoy B, Wolf SM, LawSeq Liability Task Force. From genetics to genomics: facing the liability implications in clinical care. J Law Med Ethics. 2020; 48: 11-43.
3. Pergament D, Ilijic K. The legal past, present and future of prenatal genetic testing: professional liability and other legal challenges affecting patient access to services. J Clin Med. 2014; 3: 1437-65.
4. Amagwula T, Chang PL, Hossain A, Tyner J, Rivers AL, Phelps JY. Preimplantation genetic diagnosis: a systematic review of litigation in the face of new technology. Fertil Steril. 2012; 98: 1277-82.
5. Moutos CP, Phelps JY. Wrongful birth and wrongful life lawsuits in obstetrics and gynecology. Am J Obstet Gynecol. 2024; 231: 611-7.
6. Middleton A, Patch C, Roberts J, Milne R, Costa A, Robarts L, et al. Professional duties are now considered legal duties of care within genomic medicine. Eur J Hum Genet. 2020; 28: 1301-4.
7. Vural B, Atlıoğlu E, Özbek U, Büyükdevrim S, Kolusayın Ö, Özçelik T. Introduction of DNA-based paternity determinations in Turkey. The Bulletin of Legal Medicine. 2002; 7: 5-13.
8. Uğurlubay GAA. Forensic DNA analysis in the Turkish Criminal Procedure law in the Context of the Law of Germany, Switzerland and Austria. Ceza Hukuku ve Kriminoloji Dergisi. 2017; 5: 29-87.
9. Frati P, Fineschi V, Di Sanzo M, La Russa R, Scopetti M, Severi FM, et al. Preimplantation and prenatal diagnosis, wrongful birth and wrongful life: a global view of bioethical and legal controversies. Hum Reprod Update. 2017; 23: 338-57.
10. Özkan Ö, Şahinol M, Aydinoglu AU, Aydin Son Y. Reflections on Turkish Personal Data Protection Law and genetic data in focus group discussions. Nanoethics. 2022; 16: 297-312.
11. Vatanoglu-Lutz EE. Research on embryos in Turkey with ethical and legal aspects. J Turk Ger Gynecol Assoc. 2012; 13: 191-5.
12. Suter SM. GINA at 10 years: the battle over “genetic information” continues in court. J Law Biosci. 2018; 5: 495-526.
13. Laberge AM, Burke W. Duty to warn at-risk family members of genetic disease. Virtual Mentor. 2009; 11: 656-60.
14. Mitchell C, Ploem MC, Hennekam RC, Kaye J. A duty to warn relatives in clinical genetics: arguably “fair just and reasonable” in English law? Tottels J Prof Neglig. 2016; 32: 120-36.

Original Investigation | Özgün Araştırma

Gazi Med J 2026;37(4):490-495

DOI: http://dx.doi.org/10.12996/gmj.2026.4722

Retrospective Re-evaluation of Variants of Uncertain Significance Identified by Clinical Exome Sequencing in Breast Cancer Patients

Meme Kanseri Hastalarında Klinik Ekzom Dizileme ile Belirlenen Klinik Önemi Belirsiz Varyantların Retrospektif Olarak Yeniden Değerlendirilmesi

Sezen Güntekin Ergün[1,2], Feride İffet Şahin[2]

[1]Department of Medical Biology, Hacettepe University, Faculty of Medicine, Ankara, Türkiye
[2]Department of Medical Genetics, Başkent University, Faculty of Medicine, Ankara, Türkiye

ABSTRACT

Objective: Variants of uncertain significance (VUSs) remain a major challenge in clinical molecular genetics because their interpretation may change as population, computational, and disease-specific evidence accumulates. This retrospective descriptive study aimed to re-evaluate VUSs identified by clinical exome sequencing in breast cancer patients and to characterize recurrent observations and genes of potential biological interest.

Methods: Clinical exome sequencing data from 25 individuals referred for breast cancer mutation analysis were retrospectively re-evaluated. The study population comprised individuals with breast cancer or a first-degree family history of breast cancer. Sequencing was performed on the Illumina NextSeq 550 platform. Thirteen individuals were analyzed using Sophia Clinical Exome Solution v2 with Sophia DDM, and 12 were analyzed using the Berry Genomics Clinical Exome workflow with Verita Trekker Variants Detection System and/or the Enliven Variants Annotation Interpretation System. An exploratory filtering strategy was used to focus on VUS observations outside genes already prioritized in the literature for breast cancer or cancer. Candidate variants were subsequently reviewed using public variant resources and *in silico* evidence.

Results: After sequential filtering and manual re-evaluation, four candidate observations in the once-observed group and seven candidate observations in the twice-observed group were examined in detail. Variants were identified in *ZDHHC15, COL17A1, MFN1,*

ÖZ

Amaç: Klinik önemi belirsiz varyantlar (VUS), popülasyon, hesaplamalı analiz ve hastalığa özgü kanıtların zaman içerisinde birikmesiyle yorumlarının değişebilmesi nedeniyle klinik moleküler genetikte önemli bir sorun olmaya devam etmektedir. Bu retrospektif tanımlayıcı çalışmada, meme kanseri hastalarında klinik ekzom dizileme ile belirlenen VUS'ların yeniden değerlendirilmesi ve tekrarlayan gözlemlerin yanı sıra biyolojik açıdan potansiyel ilgi taşıyan genlerin karakterize edilmesi amaçlandı.

Yöntemler: Meme kanseri mutasyon analizi amacıyla yönlendirilen 25 bireye ait klinik ekzom dizileme verileri retrospektif olarak yeniden değerlendirildi. Çalışma popülasyonu, meme kanseri tanısı bulunan veya birinci derece akrabasında meme kanseri öyküsü olan bireylerden oluşturuldu. Dizileme, Illumina NextSeq 550 platformunda gerçekleştirildi. On üç birey Sophia Clinical Exome Solution v2 ve Sophia DDM kullanılarak, 12 birey ise Berry Genomics Clinical Exome iş akışı ile Verita Trekker Variants Detection System ve/veya Enliven Variants Annotation Interpretation System kullanılarak analiz edildi. Meme kanseri veya kanser açısından literatürde daha önce önceliklendirilmiş genlerin dışındaki VUS gözlemlerine odaklanmak amacıyla keşifsel bir filtreleme stratejisi kullanıldı. Aday varyantlar daha sonra kamuya açık varyant kaynakları ve *in silico* kanıtlar kullanılarak değerlendirildi.

Bulgular: Ardışık filtreleme ve manuel yeniden değerlendirme sonrasında, bir kez gözlenen gruptaki dört aday gözlem ile iki kez gözlenen gruptaki yedi aday gözlem ayrıntılı olarak incelendi.

Cite this article as: Güntekin Ergün S, Şahin Fİ. Retrospective re-evaluation of variants of uncertain significance identified by clinical exome sequencing in breast cancer patients. Gazi Med J. 2026;37(4):490-495

Address for Correspondence/Yazışma Adresi: Feride İffet Şahin, Department of Medical Genetics, Başkent University, Faculty of Medicine, Ankara, Türkiye
E-mail / E-posta: drferidesahin@gmail.com
ORCID ID: 0000-0001-7308-9673

Received/Geliş Tarihi: 23.05.2026
Accepted/Kabul Tarihi: 27.07.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

ZNF141, MUC6, VEZT, MYBPC1, ZSWIM1, TDRD3, PRDM6, and *PIGG. MUC6* and *ZNF141* were considered of exploratory interest. Two different *MUC6* variants were observed among four individuals; however, two carriers of one *MUC6* variant were related and were not considered independent observations. Two *ZNF141* missense VUSs were identified in a single breast cancer patient. No control cohort, statistical enrichment analysis, segregation analysis, or functional validation was available.

Conclusion: The findings represent descriptive and hypothesis-generating observations rather than evidence of a causal or statistical association with breast cancer. *MUC6* and *ZNF141* may warrant further investigation in larger, population-matched cohorts with segregation, gene-level enrichment, and functional studies. The study also emphasizes the importance of transparent and periodic reinterpretation of VUSs.

Keywords: Breast cancer, variant of uncertain significance, VUS, clinical exome sequencing, variant reinterpretation

ÖZ

ZDHHC15, COL17A1, *MFN1*, *ZNF141, MUC6, VEZT, MYBPC1, ZSWIM1, TDRD3*, *PRDM6* ve *PIGG* genlerinde varyantlar saptandı. *MUC6* ve *ZNF141* keşifsel açıdan ilgi çekici bulundu. Dört bireyde iki farklı *MUC6* varyantı gözlendi; ancak *MUC6* varyantlarından birini taşıyan iki bireyin akraba olduğu ve bu nedenle bağımsız gözlemler olarak değerlendirilmediği belirlendi. Tek bir meme kanseri hastasında iki farklı *ZNF141* missense VUS saptandı. Kontrol kohortu, istatistiksel zenginleşme analizi, segregasyon analizi veya fonksiyonel doğrulama verisi bulunmamaktaydı.

Sonuç: Bulgular, meme kanseri ile nedensel veya istatistiksel bir ilişkiyi ortaya koyan kanıtlardan ziyade tanımlayıcı ve hipotez oluşturucu gözlemler niteliğindedir. *MUC6* ve *ZNF141* genleri; segregasyon, gen düzeyinde zenginleşme ve fonksiyonel çalışmaların gerçekleştirileceği, daha geniş ve popülasyon açısından eşleştirilmiş kohortlarda ileri araştırmaları hak edebilir. Çalışma ayrıca VUS'ların şeffaf ve düzenli aralıklarla yeniden yorumlanmasının önemini vurgulamaktadır.

Anahtar Sözcükler: Meme kanseri, klinik önemi belirsiz varyant, VUS, klinik ekzom dizileme, varyantın yeniden yorumlanması

INTRODUCTION

Cancer is a multifactorial disease characterized by disruption of cellular homeostasis and uncontrolled cell proliferation resulting from the interaction of genetic and environmental factors (1). Breast cancer is a heterogeneous malignancy arising predominantly from epithelial cells of the mammary ducts or lobules (2). Its development is influenced by age, hormonal and reproductive factors, lifestyle, breast density, environmental exposures, and family history (3).

Although breast cancer is a genetic disease at the cellular level, it is not always hereditary. *BRCA1* and *BRCA2* are the best-established high-penetrance hereditary breast and ovarian cancer genes. Additional genes, including *ATM, CHEK2, PALB2, TP53, PTEN, CDH1*, and *STK11*, contribute to breast cancer susceptibility with variable penetrance (4-6). The increasing use of next-generation sequencing and clinical exome sequencing has expanded the number of variants detected during genetic evaluation and has consequently increased the number of variants of uncertain significance (VUSs) requiring interpretation.

A VUS is a sequence variant for which the available evidence is insufficient or conflicting for classification as pathogenic/likely pathogenic or benign/likely benign. A VUS should not be considered evidence of causality or used alone to establish a clinical association. Because population databases, computational predictions, clinical submissions, and functional evidence evolve over time, periodic reinterpretation of previously identified VUSs may provide additional information.

The aim of the present study was to retrospectively re-evaluate VUSs identified by clinical exome sequencing in breast cancer patients and to characterize recurrent observations and genes of potential biological interest. The study was designed as a descriptive and exploratory reanalysis and was not intended to establish statistical gene–disease associations.

MATERIALS AND METHODS

Study Design and Breast Cancer Patient Cohort

This retrospective descriptive study included 25 individuals referred to the Department of Medical Genetics, Başkent University Hospitals, for breast cancer mutation analysis between December 2019 and October 2021. The study population comprised individuals with a diagnosis of breast cancer or a first-degree family history of breast cancer. Throughout the manuscript, the cohort is referred to in the context of breast cancer patients and breast cancer genetic evaluation. All included individuals had previously undergone clinical exome sequencing as part of their genetic assessment.

The study was approved by the Başkent University Medical and Health Sciences Research Board on April 12, 2022 (decision no. 119004; project no. KA22/45) and was supported by the Başkent University Research Fund. Written informed consent was obtained from all participants.

Clinical Exome Sequencing and Primary Analysis

Clinical exome sequencing was performed using the Illumina NextSeq 550 platform. Thirteen individuals were analyzed using the Sophia Clinical Exome Solution v2 assay, and primary analysis was performed using Sophia DDM software. The remaining 12 individuals were analyzed using the Berry Genomics Clinical Exome workflow. Variant detection and annotation for the Berry Genomics data were performed using the Verita Trekker Variants Detection System and the Enliven Variants Annotation Interpretation System. The gene contents of the Sophia Clinical Exome Solution v2 and Berry Genomics Clinical Exome assays are provided in Supplementary S1 and S2, respectively.

The present study used retrospectively available variant outputs generated by these clinical workflows. Because two clinical exome workflows and different output structures were used during the

study period, the datasets were standardized before exploratory filtering. The Sophia-derived dataset initially contained 101,225 records. The Berry Genomics outputs contained 195,241 records from the Verita Trekker Variants Detection System and 49,751 records from the Enliven Variants Annotation Interpretation System; because the latter two datasets had similar output characteristics, they were combined for subsequent evaluation.

Exploratory VUS Reanalysis and Filtering Strategy

The objective of the exploratory reanalysis was to examine VUS observations outside genes already prioritized in breast cancer and general cancer literature. At the first stage, a literature-derived list of 2,156 cancer-related genes was used as an exclusion list. This strategy was not intended to imply that known cancer genes are biologically unimportant; rather, it was used to focus the exploratory analysis on observations that would not routinely be prioritized as established cancer-gene findings. The complete list of excluded genes is provided as Supplementary S3.

After application of the 2,156-gene exclusion list, 64,677 records remained in the Sophia dataset and 165,881 records remained in the combined Berry Genomics dataset. Known benign and pathogenic variants were then removed to enrich the dataset for variants reported as VUSs or variants requiring further review. This sequential process yielded 38,173 records for additional evaluation.

The original filtering description did not adequately distinguish automated consequence filtering from the subsequent manual and *in silico* review. In the revised workflow, non-coding and synonymous observations were deprioritized during the initial consequence-based screen; however, recurrent records that remained unresolved or inconsistently classified by the source analysis systems were manually reviewed. Therefore, selected intronic observations were retained for descriptive evaluation. Their retention did not constitute evidence of pathogenicity, and intronic or synonymous observations were not interpreted as disease-associated in the absence of splicing, RNA, or other functional evidence.

Using record counts across the standardized datasets, 14 variants were initially identified in the once-observed group and 868 in the twice-observed group. These variants were individually re-evaluated using *in silico* and public variant interpretation resources. Variants that could be resolved as benign during this review were deprioritized. Four candidate observations from the once-observed group and seven candidate observations from the twice-observed group were retained for detailed descriptive assessment. In this manuscript, “once-observed” and “twice-observed” refer to occurrence counts within the standardized retrospective dataset and should not be interpreted as population frequencies or statistical recurrence.

Variant Interpretation and Database Review

Candidate variants were reviewed using Franklin, VarSome, and gnomAD, together with available literature and gene-level biological information. ACMG/AMP evidence codes were treated as evidence criteria rather than numerical scores. In accordance with current ClinGen Sequence Variant Interpretation guidance, rarity/absence evidence and computational or splicing evidence were interpreted cautiously and in the context of the relevant gene–disease relationship. PM2, BP4, and BP7 codes reported by the source databases are presented as database-reported evidence criteria; they were not independently counted as proof of breast cancer association. BP4 and BP7 were explicitly recognized as benign-supporting evidence. Because several candidate genes do not have an established breast cancer gene–disease relationship and transcript-specific raw evidence was not uniformly available for retrospective recuration, we did not claim a *de novo* definitive ACMG/AMP classification for every variant. Instead, the revised tables distinguish database-reported ACMG/AMP evidence, public database classification, and the interpretation adopted in the present reanalysis.

Clinical Variables and Analytical Approach

The cohort comprised 25 individuals aged 25-64 years. The retrospective clinical records identified 18 individuals with a breast cancer diagnosis, including one individual with fibroadenomatous hyperplasia and breast cancer, and 7 individuals referred primarily because of family history.

Because the total cohort size was small and no healthy, population-matched, or internal control cohort was available, no inferential gene-level burden or enrichment analysis was performed. Related individuals were not treated as independent observations in frequency-based interpretation. Available clinical characteristics were summarized descriptively. Molecular subtype, receptor status, tumor stage, and tumor grade were not consistently available in the retrospective dataset and therefore could not be used for adequately powered subgroup analyses.

Statistical Analysis

Statistical analysis was descriptive. Categorical clinical variables (diagnostic category and referral indication) were summarized as counts, age as a range, and variant-level results as record counts at each filtering step; “once-observed” and “twice-observed” denote occurrence counts in the standardized retrospective dataset, not allele frequencies. Related individuals were not counted as independent recurrence events. No inferential analysis (hypothesis testing, p-values, confidence intervals, odds ratios, gene-level burden or enrichment tests, or multiple-testing correction) was performed, because no control cohort was available, the sample size was small, and the study was not powered for association testing; accordingly, no finding should be interpreted as statistically significant. Clinical variables that were not consistently available (molecular subtype, receptor status, tumor stage, and tumor grade) were neither analyzed nor imputed.

RESULTS

A total of 25 individuals referred for breast cancer mutation analysis were included in the retrospective reanalysis. The cohort was evaluated as an exploratory breast cancer series rather than as a case-control association cohort. At the beginning of the bioinformatic review, variant records were evaluated independently of the individual clinical characteristics; clinical context was considered only after the filtering and candidate-review stages.

After application of the literature-derived 2,156-gene exclusion list and removal of known benign and pathogenic variants, 38,173 records remained for further evaluation. The initial occurrence-

based review identified 14 variants in the once-observed group and 868 variants in the twice-observed group. Individual *in silico* and database review further reduced these groups to four and seven candidate observations, respectively.

The detailed once-observed candidate group involved *ZDHHC15, COL17A1, MFN1*, and *ZNF141*. The *ZDHHC15* c.137-51_137-50delGA, *COL17A1* c.1688-45_1688-44dupCA, and *MFN1* c.1663-8_1663-6dupTTT variants were intronic observations. Public interpretation resources included benign-supporting evidence for these variants, and *COL17A1* and *MFN1* were classified as benign or likely benign in at least one reviewed resource. Accordingly, these observations were retained only for descriptive completeness and were not interpreted as evidence of breast cancer susceptibility (Table 1).

Two *ZNF141* missense variants, c.136G>A (p.Ala46Thr) and c.193G>T (p.Val65Leu), were identified in the same 46-year-old female breast cancer patient. Both variants were reported as VUSs in the reviewed resources and had BP4 benign-supporting computational evidence. The patient also had previously reported VUSs in *ERBB2, PMS1,* and *TSC1*. The co-occurrence of two *ZNF141* VUSs in one patient was considered an exploratory observation and not evidence of a *ZNF141*–breast cancer association.

The detailed twice-observed candidate group involved *MUC6, VEZT, MYBPC1, ZSWIM1, TDRD3, PRDM6*, and *PIGG*. Several of these observations were synonymous or intronic and were therefore interpreted cautiously. Gene-level literature suggesting roles in cancer-related biological processes was not considered variant-level evidence of pathogenicity.

Two different *MUC6* variants were identified among four individuals. One *MUC6* variant was observed in two unrelated individuals aged 25 and 49 years; the 25-year-old individual had fibroadenomatous hyperplasia and breast cancer, whereas the 49-year-old individual was referred in the context of a family history of breast cancer. The other *MUC6* variant was observed in two sisters aged 43 and 45 years from a family with breast cancer history. Because the sisters were related, they were not considered independent recurrence events. No formal segregation analysis involving affected and unaffected relatives was performed (Table 2).

No control-based comparison, population-matched enrichment test, gene-level burden analysis, formal segregation study, or functional validation was performed. Therefore, the present data do not demonstrate a statistical or causal association between the identified variants or genes and breast cancer.

DISCUSSION

In this retrospective descriptive study, clinical exome sequencing data from breast cancer patients were re-evaluated with a specific focus on VUS observations outside genes already prioritized in cancer literature. The study identified a limited set of recurrent or potentially noteworthy observations after sequential filtering and manual review. The findings should be interpreted as hypothesis-generating and do not establish gene–disease associations.

The original exploratory strategy excluded a literature-derived list of 2,156 genes associated with breast cancer or cancer in general. We acknowledge that excluding established cancer genes would be inappropriate in a comprehensive clinical diagnostic analysis. In the present study, however, the strategy was used only after the original clinical exome evaluation and was intended to focus a secondary research reanalysis on less routinely prioritized genes. The known cancer-related findings from the original clinical reports were not considered absent or clinically irrelevant; rather, they were outside the specific exploratory question of this secondary analysis.

MUC6 was considered of exploratory interest because two different variants were observed among four individuals. However, the interpretation is limited by family relatedness, the absence of segregation analysis, and the absence of a control population. *MUC6* encodes a secreted mucin involved in epithelial protection and has been studied as a biomarker in several epithelial malignancies (7-9). In breast cancer, mucin expression studies have reported variation in *MUC6* expression according to tumor characteristics, including estrogen receptor status and mucinous histology (10). These gene-

Table 1. List of recurring variants.

Gene	Nucleotide variation	Protein variation	dbSNP	Consequence	ACMG/AMP evidence reported in source databases	Public database classification	Interpretation in the present reanalysis
ZDHHC15	c.137-51_137-50delGA	—	rs773803900	Intronic	BP7	VUS	Descriptive only; benign-supporting evidence
COL17A1	c.1688-45_1688-44dupCA	—	rs372506043	Intronic	PM2, BP7	Benign	Descriptive only; conflicting/benign evidence
MFN1	c.1663-8_1663-6dupTTT	—	rs75719342	Intronic	BP7	Likely benign	Deprioritized; benign-supporting evidence
ZNF141	c.136G>A	p.(Ala46Thr)	rs75245145	Missense	BP4; PM2 reported by VarSome	VUS	VUS; hypothesis-generating only
ZNF141	c.193G>T	p.(Val65Leu)	rs78725891	Missense	BP4; PM2 reported by VarSome	VUS	VUS; hypothesis-generating only

ACMG/AMP codes are evidence criteria, not numerical scores. PM2 is pathogenic-supporting rarity/absence evidence when appropriately applicable; BP4 and BP7 are benign-supporting evidence. Criteria shown here reflect evidence reported by the source databases used in the retrospective analysis and were not treated as independent proof of breast cancer association. VUS: Variant of uncertain significance, —: Not available/not applicable.

Table 2. Candidate variants in the twice-observed group.

Gene	Nucleotide variation	Protein variation	dbSNP	Consequence	ACMG/AMP evidence reported in source databases	Public database classification	Interpretation in the present reanalysis
MUC6	c.5T>G	p.(Val2Gly)	rs201440145	Missense	PM2 reported	VUS	VUS; hypothesis-generating only
MUC6	c.4967C>T	p.(Pro1656Leu)	—	Missense	PM2, BP4 reported	VUS	VUS; hypothesis-generating only
VEZT	c.634delC	p.(Arg212GlufsTer8)	—	Frameshift	PM2 reported	VUS	VUS; no PVS1 assigned without gene/disease-specific LoF assessment
MYBPC1	c.1793T>A	p.(Ile598Lys)	—	Missense	PM2 reported	VUS	VUS; no breast-cancer association inferred
ZSWIM1	c.627C>T	p.(Asn209=)	—	Synonymous	PM2, BP7 reported	Not available	Descriptive only; BP7 is benign-supporting
TDRD3	c.1019G>A	p.(Arg340Lys)	—	Missense	PM2 reported; BP4 reported by VarSome	VUS	VUS; hypothesis-generating only
PRDM6	c.561G>A	p.(Pro187=)	—	Synonymous	PM2, BP7 reported	Likely benign	Deprioritized; benign-supporting evidence
PIGG	c.2871C>T	p.(Tyr957=)	—	Synonymous	PM2, BP7 reported; BP4/BP7 by VarSome	Likely benign	Deprioritized; benign-supporting evidence

ACMG/AMP codes are evidence criteria, not numerical scores. PM2 is pathogenic-supporting rarity/absence evidence when appropriately applicable; BP4 and BP7 are benign-supporting evidence. Criteria shown here reflect evidence reported by the source databases used in the retrospective analysis and were not treated as independent proof of breast cancer association. VUS: Variant of uncertain significance, —: Not available/not applicable.

level observations provide biological context but do not demonstrate that the *MUC6* variants identified in our cohort confer breast cancer susceptibility.

ZNF141 was also of exploratory interest because two missense VUSs were observed in a single breast cancer patient. Zinc-finger proteins have diverse regulatory functions, and several members of the broader zinc-finger protein family have been implicated in cancer biology (11-16). Nevertheless, evidence concerning other zinc-finger genes cannot be extrapolated to *ZNF141*. Moreover, both *ZNF141* variants had BP4 benign-supporting computational evidence. The co-occurrence of two VUSs in one individual is insufficient to infer causality, and larger datasets and gene-specific functional studies are required.

The intronic observations in *ZDHHC15, COL17A1*, and *MFN1* illustrate an important limitation of retrospective VUS review. Although such variants may remain unresolved in one resource or source workflow, benign-supporting evidence, location outside canonical splice regions, or benign/likely benign classifications in other resources substantially weaken disease-related interpretation. These variants were therefore discussed descriptively and were not used to support breast cancer association. The same caution applies to synonymous observations: gene-level cancer literature does not establish a functional effect for a specific synonymous variant.

The ACMG/AMP framework is an evidence-based classification system rather than a numerical scoring system (17). PM2, BP4, and BP7 are evidence criteria with different directions and strengths. Current ClinGen guidance emphasizes transparent, criterion-specific application of rarity, computational, and splicing evidence. In particular, BP4 and BP7 are benign-supporting criteria and cannot be used to support disease relevance. We therefore did not convert automated Franklin or VarSome outputs into independent evidence of breast cancer association. The revised tables separately report database-reported ACMG/AMP evidence, public database classification, and our interpretation in this retrospective reanalysis. A definitive *de novo* ACMG/AMP classification was not assigned where the required gene–disease validity, transcript-specific, segregation, or functional evidence was unavailable.

Breast cancer is clinically and molecularly heterogeneous. The reviewer appropriately suggested analyses by age at diagnosis, family history, histopathological subtype, molecular subtype, stage, grade, and *BRCA1/BRCA2* status. Age and family history were available and are now presented descriptively. However, receptor status, molecular subtype, stage, and grade were not consistently available in the retrospective source dataset, and the cohort size of 25 individuals would result in very small subgroup counts. We therefore did not perform inferential subgroup analyses that could yield unstable or misleading estimates.

Periodic reinterpretation of VUSs remains important because population resources, clinical databases, computational evidence, and functional data evolve. Nevertheless, reinterpretation should not be conflated with association testing. Robust gene–disease association requires appropriate comparison cohorts, population matching, independent observations, statistical enrichment or burden analyses, and preferably segregation and functional evidence. The present study provides candidate observations that may inform future research under these more rigorous designs.

Study Limitations

This study has several limitations. The sample size was small and the retrospective cohort was not designed or powered for variant-level

or gene-level association testing. No healthy, population-matched, or internal control cohort was available, precluding enrichment and burden analyses. The cohort was clinically heterogeneous, and molecular subtype, receptor status, tumor stage, and tumor grade were not consistently available for subgroup analysis. Two clinical exome workflows with different output structures were used, requiring retrospective standardization of variant records. The 2,156-gene exclusion strategy was an exploratory research filter and was not a validated disease-association pipeline. Related individuals were present among the *MUC6* observations and no formal segregation analysis was performed. No RNA, splicing, or other functional validation was available. Finally, VUS classifications and database evidence evolve over time. These limitations preclude causal or statistical association claims.

CONCLUSION

In conclusion, this retrospective descriptive reanalysis of clinical exome data from breast cancer patients identified a limited number of exploratory VUS observations after sequential filtering and database review. *MUC6* and *ZNF141* were of potential biological interest; however, the current data do not demonstrate that variants in these genes are associated with breast cancer. The observations should be considered hypothesis-generating and require evaluation in larger, population-matched cohorts incorporating independent recurrence, segregation analysis, gene-level statistical testing, and functional validation. The study also emphasizes the importance of cautious, transparent, and periodic reinterpretation of VUSs in clinical molecular genetics.

Ethics

Ethics Committee Approval: The study was approved by the Başkent University Medical and Health Sciences Research Board on April 12, 2022 (decision no. 119004; project no. KA22/45) and was supported by the Başkent University Research Fund.

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

Footnotes

Authorship Contributions

Surgical and Medical Practices: F.İ.Ş., Concept: S.G.E., F.İ.Ş., Design: S.G.E., F.İ.Ş., Data Collection or Processing: S.G.E., Analysis or Interpretation: S.G.E., Literature Search: S.G.E., Writing S.G.E., F.İ.Ş.

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

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

Supplement 1 Link: https://d2v96fxpocvxx.cloudfront.net/cf9d60d6-523c-458a-a2e6-78728d3ffbb0/content-images/4ddd9c81-7e43-453e-a8fa-1a48305e1313.pdf

Supplement 2 Link: https://d2v96fxpocvxx.cloudfront.net/cf9d60d6-523c-458a-a2e6-78728d3ffbb0/content-images/11b3ce4a-8b7b-4def-b8a5-d5ba070331a2.pdf

Supplement 3 Link: https://d2v96fxpocvxx.cloudfront.net/cf9d60d6-523c-458a-a2e6-78728d3ffbb0/content-images/e6702b2d-792d-4a16-a02f-d45a656a92cd.pdf

REFERENCES

1. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022; 12: 31-46.
2. Carbone PP, Jordan VC, Bonadonna G. Neoplasms of the breast. In: Calabresi P, Schein PS, editors. Medical oncology: basic principles and clinical management of cancer. 2nd ed. New York: McGraw-Hill; 1993. p. 819-49.
3. Lei S, Zheng R, Zhang S, Wang S, Chen R, Sun K, et al. Global patterns of breast cancer incidence and mortality: a population-based cancer registry data analysis from 2000 to 2020. Cancer Commun. 2021; 41: 1183- 94.
4. Shih HA, Nathanson KL, Seal S, Collins N, Stratton MR, Rebbeck TR, et al. BRCA1 and BRCA2 mutations in breast cancer families with multiple primary cancers. Clin Cancer Res. 2000; 6: 4259-64.
5. Foulkes WD. The ten genes for breast (and ovarian) cancer susceptibility. Nat Rev Clin Oncol. 2021;18: 259-60.
6. Breast Cancer Association Consortium; Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C, Wahlström C, et al. Breast cancer risk genes - association analysis in more than 113,000 women. N Engl J Med. 2021; 384: 428-39.
7. Bartman AE, Buisine MP, Aubert JP, Niehans GA, Toribara NW, Kim YS, et al. The MUC6 secretory mucin gene is expressed in a wide variety of epithelial tissues. J Pathol. 1998; 186: 398-405.
8. Yuki A, Fujii C, Yamanoi K, Matoba H, Harumiya S, Kawakubo M, et al. Glycosylation of MUC6 by α1,4-linked N-acetylglucosamine enhances suppression of pancreatic cancer malignancy. Cancer Sci. 2022; 113: 576-86.
9. Yamanoi K, Fujii C, Yuzuriha H, Kumazawa M, Shimoda M, Emoto K, et al. MUC6 expression is a preferable prognostic marker for invasive mucinous adenocarcinoma of the lung. Histochem Cell Biol. 2022; 157: 671-84.
10. Rakha EA, Boyce RW, Abd El-Rehim D, Kurien T, Green AR, Paish EC, et al. Expression of mucins (MUC1, MUC2, MUC3, MUC4, MUC5AC and MUC6) and their prognostic significance in human breast cancer. Mod Pathol. 2005; 18: 1295-304.
11. Jen J, Wang YC. Zinc finger proteins in cancer progression. J Biomed Sci. 2016; 23: 53.
12. Yang H, Lu Y, Zheng Y, Yu X, Xia X, He X, et al. shRNA-mediated silencing of ZFX attenuated the proliferation of breast cancer cells. Cancer Chemother Pharmacol. 2014; 73: 569-76.
13. Cheng Y, Liang P, Geng H, Wang Z, Li L, Cheng SH, et al. A novel 19q13 nucleolar zinc finger protein suppresses tumor cell growth through inhibiting ribosome biogenesis and inducing apoptosis but is frequently silenced in multiple carcinomas. Mol Cancer Res. 2012; 10: 925-36.
14. Harper J, Yan L, Loureiro RM, Wu I, Fang J, D'Amore PA, et al. Repression of vascular endothelial growth factor expression by the zinc finger transcription factor ZNF24. Cancer Res. 2007; 67: 8736-41.
15. Hu R, Peng G, Dai H, Breuer EK, Stemke-Hale K, Li K, et al. ZNF668 functions as a tumor suppressor by regulating p53 stability and function in breast cancer. Cancer Res. 2011; 71: 6524-34.
16. Bassey-Archibong BI, Kwiecien JM, Milosavljevic SB, Hallett RM, Rayner LG, Erb MJ, et al. Kaiso depletion attenuates transforming growth factor-β signaling and metastatic activity of triple-negative breast cancer cells. Oncogenesis. 2016; 5: e208.
17. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al.; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015; 17: 405-24.

Gazi Med J 2026;37(4):496-503

DOI: http://dx.doi.org/10.12996/gmj.2026.4729

Study Habits and Attitudes Toward Artificial Intelligence Tools Among Anesthesia Trainees in Türkiye: A Descriptive Survey

Türkiye'deki Anestezi Asistanlarında Çalışma Alışkanlıkları ve Yapay Zeka Araçlarına Yönelik Tutumlar: Tanımlayıcı Bir Anket Çalışması

Çağrı Özdemir[1], Beyza Büyükgebiz Yeşil[2], Miray Gözde Özdemir[3]

[1]Department of Anesthesiology and Reanimation, Gazi University, Faculty of Medicine, Ankara, Türkiye
[2]Clinic of Anesthesiology and Reanimation, Kahramanmaraş State Hospital, Kahramanmaraş, Türkiye
[3]Department of Anesthesiology and Reanimation, University of Health Sciences Türkiye, Ankara Etlik City Hospital, Ankara, Türkiye

ABSTRACT

Objective: To describe the study habits and attitudes toward the use of artificial intelligence (AI) tools for studying among anesthesia trainees in Türkiye, and to explore the associations between these factors and performance in national in-training examinations.

Methods: A cross-sectional online survey was administered to anesthesia trainees in Türkiye between April and June 2025. The questionnaire addressed study frequency and duration, preferred resources, AI use patterns, and 5-point Likert ratings for reliability and usefulness. Performance in the national online trainee assessment examination (ÇADS) was analyzed using Kruskal-Wallis tests and Spearman correlation, with ÇADS scores encoded as ordinal ranks (1-4).

Results: A total of 148 responses were analyzed. Study sessions were reported as "occasional" by 60.1%, with typical sessions lasting 1-2 hours (48.6%). Although 96.6% of the respondents perceived a need to study more, the main barriers were fatigue and burnout (46.9%) and time constraints due to clinical duties (29.4%). The most commonly used resources were textbooks (92.6%), spaced repetition (60.8%), and online journals (52.7%). AI tools were rated as useful (mean 3.89 ± 0.99) but only moderately reliable (3.22 ± 0.97). Two-thirds of the respondents participated regularly in ÇADS (66.2%). Among the 130 participants who had completed at least one ÇADS examination, the most frequently reported score category was 60-79 (73.1%). Frequent AI use was significantly associated with higher ÇADS performance (p =

ÖZ

Amaç: Türkiye'de anesteziyoloji ve reanimasyon asistanlarının klinik dışı çalışma alışkanlıklarını, kullandıkları eğitim kaynaklarını ve yapay zeka (YZ) tabanlı araçlara ilişkin kullanım sıklığını ortaya koymak, ayrıca bu faktörlerin ulusal çevrim içi asistan değerlendirme sınavı (ÇADS) performansıyla ilişkisini değerlendirmektir.

Yöntemler: Nisan-Haziran 2025 tarihleri arasında anesteziyoloji asistanlarına çevrimiçi, anonim, kesitsel bir anket uygulandı. Anket; demografik özellikler, çalışma sıklığı ve süresi, eğitim kaynakları, YZ kullanım sıklığı ve 5'li Likert ölçeği ile güvenilirlik/fayda algısı sorularından oluştu. ÇADS skorları sıralı kategoriler (1-4) olarak kodlandı ve Kruskal-Wallis testi ve Spearman korelasyonu ile değerlendirildi.

Bulgular: Toplam 148 yanıt analiz edildi. Çalışma sıklığı ağırlıklı olarak "ara sıra" (%60,1), oturum süresi 1-2 saat (%48,6) olarak bildirildi. Katılımcıların %96,6'sı daha fazla çalışması gerektiğini düşünüyordu; ana engeller yorgunluk/tükenmişlik (%46,9) ve klinik görevlerden kaynaklı zaman eksikliği (%29,4) idi. En sık kullanılan kaynaklar ders kitapları (%92,6), aralıklı tekrar (%60,8) ve çevrimiçi makaleler (%52,7) iken YZ araçlarını düzenli kullananlar yalnızca %24,3 oranındaydı. YZ'nin faydalılığı ortalama 3,89 ± 0,99, güvenilirliği 3,22 ± 0,97 olarak puanlandı. ÇADS sınavına katılan 130 asistanın %73,1'inin son ÇADS puanı 60-79 aralığındaydı. Düzenli YZ kullanımı (günlük veya haftada 2-3 kez) ÇADS performansıyla anlamlı ilişkiliydi (p = 0,003). Kendine güven düzeyi de ÇADS performansıyla pozitif korele idi (r_s = 0,205, p = 0,019). Çalışma sıklığı ve kullanılan teknikler ÇADS performansıyla ilişkili değildi.

Cite this article as: Özdemir Ç, Büyükgebiz Yeşil B, Özdemir MG. Study habits and attitudes toward artificial intelligence tools among anesthesia trainees in Türkiye: a descriptive survey. Gazi Med J. 2026;37(4):496-503

Address for Correspondence/Yazışma Adresi: Çağrı Özdemir, Department of Anesthesiology and Reanimation, Gazi University, Faculty of Medicine, Ankara, Türkiye
E-mail / E-posta: cagriozdemir@gazi.edu.tr
ORCID ID: orcid.org/0000-0002-1266-8054

Received/Geliş Tarihi: 02.06.2026
Accepted/Kabul Tarihi: 17.08.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

0.003). No associations were found between ÇADS performance and study frequency or specific study techniques.

Conclusion: Anesthesia trainees in Türkiye display irregular, brief study patterns, with fatigue and burnout as the main problems. AI tool adoption remains low, likely due to moderate trust and a lack of guidance. Frequent AI use and higher self-confidence were associated with better examination performance in exploratory univariable analyses. Educational programs may benefit from addressing workload and burnout while providing structured guidance for the appropriate integration of AI tools.

Keywords: Anesthesia trainees, study habits, artificial intelligence, e-learning, Türkiye

ÖZ

Sonuç: Türkiye'deki anesteziyoloji asistanları düzensiz ve kısa süreli çalışma örüntüsü sergilemekte; yorgunluk ve tükenmişlik temel sorunlar olarak öne çıkmaktadır. YZ araçları düşük güven ve rehberlik eksikliği nedeniyle az kullanılmaktadır. Tek değişkenli keşifsel analizlerde, sık YZ kullanımı ve yüksek kendine güven düzeyinin daha iyi sınav performansı ile ilişkili olduğu bulundu. Eğitim programları iş yükü ve tükenmişliği doğrudan ele almalı, YZ araçlarının entegrasyonu desteklenmelidir.

Anahtar Sözcükler: Anesteziyoloji, çalışma alışkanlıkları, yapay zeka, çevrimiçi öğrenme, Türkiye

INTRODUCTION

In medical education, trainee study behavior is not uniform across specialties, institutions, or even training years; it is an adaptive response to workload, assessment pressure, and the availability of structured learning tools.

Beyond bedside learning, trainees are expected to engage in self-directed study to maintain clinical competence and prepare for in-training or board examinations. Accordingly, study habits constitute an important component of residency education, as they shape not only examination preparation but also the development of long-term learning routines (1-4).

Previous studies from different specialties suggest that trainees' study behaviors are influenced by multiple factors, including workload, duty-hour structure, educational resources, and the assessment system within their training program (1-5). Trainees often rely on a combination of traditional methods, such as textbooks and written summaries, and more structured or exam-oriented approaches, including case conferences, question-based learning, and online modules (1-3). Access to digital resources and structured educational formats has also been associated with more consistent study patterns and, in some reports, with improved examination performance (1,2,5). At the same time, external pressures such as heavy service demands, fatigue, and disrupted clinical environments may reduce the time and energy available for self-study (5-7).

These issues may be particularly relevant in anesthesiology, a specialty characterized by intense clinical responsibility, irregular working hours, night shifts, and a high cognitive load. Although anesthesiology trainees are expected to master broad theoretical knowledge while simultaneously developing procedural and decision-making skills, little is known about how they actually study outside clinical hours, which learning resources they prefer, and what barriers limit their study routines. This gap is important because educational planning is often based on assumptions rather than on direct data from the trainees themselves.

Since their public release in November 2022, artificial intelligence (AI) chatbots (also known as large language models or LLMs) have attracted substantial interest in medical education as fast and user-friendly tools that may support students, trainees, educators, and health care professionals in accessing and processing information (8-10). Early reports have shown that LLMs can achieve passing or near-passing performance on components of medical examinations, including the United States Medical Licensing Examination, highlighting their potential relevance for knowledge acquisition and assessment (11,12). In addition to retrieving and summarizing information, LLMs may assist learners by generating explanations, clarifying difficult concepts, and supporting question-based learning in an interactive and low-stakes environment (9,10,12). Such tools may also facilitate academic tasks by providing rapidly synthesized information that can be used for studying and writing purposes (13). However, despite increasing interest in AI-assisted learning, the actual role of these tools in postgraduate medical education and their integration into trainees' everyday study routines remain insufficiently defined.

Despite growing interest in both trainee study behavior and digital learning tools, data from Türkiye remain limited. Findings from other countries may not be directly generalizable because residency training conditions vary substantially across health-care systems in terms of workload, institutional structure, educational culture, and access to learning resources. Understanding how anesthesiology trainees in Türkiye study, which resources they use, and how they perceive AI-based tools may provide useful insights for training programs seeking to improve learning support under real-world clinical constraints.

The present study aimed to evaluate the study habits of anesthesiology trainees in Türkiye, identify the main barriers affecting self-directed learning, describe the educational resources they use, and assess their attitudes toward the use of AI tools for studying and examination preparation.

MATERIALS AND METHODS

This was a national, cross-sectional, web-based survey conducted to explore the habits, preferred learning resources, and attitudes toward AI tools of anesthesiology trainees in Türkiye. Ethical approval was obtained from the Gazi University Ethics Commission (approval no: 13.01.2025-E.1141829, date: 13.01.2025). Participation was voluntary and anonymous, and all respondents provided electronic informed consent before proceeding to the survey. No financial or material incentives were offered. The study was reported in accordance with the Checklist for Reporting Results of Internet E-Surveys (14).

The questionnaire was developed by the authors specifically for this study to examine how anesthesiology trainees study outside clinical working hours and how they perceive emerging AI-based educational tools. After reviewing the relevant literature on resident study habits, examination preparation, digital learning resources, and AI-assisted education, the initial item pool was reviewed internally by the study team for relevance, clarity, and content coverage before pilot testing. Before distribution, the survey was pilot-tested among a small group of anesthesiology trainees who were not included in the final analysis, and minor wording revisions were made to improve clarity and completion flow. Although the questionnaire underwent internal review and pilot testing, it was not formally validated using psychometric methods; therefore, measurement error cannot be excluded. The final survey included items addressing demographic and training-related characteristics, frequency and duration of study outside clinical hours, commonly used study methods and educational resources, and patterns of AI tool use for learning and examination preparation. Additional questions addressed participation in national and international examinations, including the Turkish Society of Anaesthesiology and Reanimation online trainee assessment examinations (ÇADS) and the European Diploma in Anaesthesiology and Intensive Care (EDAIC).

The survey was prepared and administered through Typeform® (Typeform S.L., Barcelona, Spain), and the survey link was circulated electronically through WhatsApp groups used by anesthesiology trainees and other professional communication networks targeting anesthesiology trainees in Türkiye. This study therefore used a convenience sample of anesthesiology trainees who voluntarily responded to the survey through these professional communication networks. Because the exact number of trainees who received or viewed the invitation could not be determined, a formal response rate could not be calculated. Data collection was carried out between April and June 2025. Only fully completed questionnaires were included in the analysis, whereas incomplete submissions were excluded. Because the survey was conducted anonymously, no directly identifying personal information was retained in the final dataset.

No formal a priori sample size calculation was performed because the study was designed as an exploratory, web-based descriptive survey. The final sample size was comparable to those reported in previous survey-based studies evaluating study habits and academic performance among medical residents. For example, Barragán-Hervella et al. (4) analyzed 112 residents using a Likert-scale questionnaire and correlation-based analyses.

Statistical Analysis

Data were analyzed using StatsDirect (version 4.0.4, StatsDirect Ltd., UK). Categorical variables were summarized as numbers and percentages. Continuous variables were presented as mean ± standard deviation (SD), whereas Likert-scale items (1-5) were summarized using both mean ± SD and median with interquartile range (IQR). For analytical purposes, ÇADS score categories were encoded as ordinal ranks (1 = < 40, 2 = 40-59, 3 = 60-79, 4 = 80-100) and presented as medians with IQRs.

For group comparisons, the Kruskal-Wallis test was used for ordinal outcomes with more than two groups, followed by Dunn's post-hoc test with Bonferroni correction where appropriate. The Mann-Whitney U test was applied for pairwise comparisons of the study techniques. Spearman's rank correlation coefficient (r_s) was calculated to assess associations between ordinal variables (self-confidence level and ÇADS performance). Categorical associations were examined using the chi-square test or Fisher's exact test where appropriate.

A two-sided p-value below 0.05 was considered statistically significant. All analyses were descriptive and exploratory in nature given the cross-sectional design.

RESULTS

A total of 148 responses were received and included in the analysis. The majority of respondents were aged 25-30 years (n = 111, 75.0%), with female participants comprising 56.8% (n = 84) of the sample. Detailed characteristics of the participants are shown in Table 1.

The majority of respondents reported studying outside clinical hours only "occasionally" (n = 89, 60.1%), while regular study (2-3 times per week or more frequently) was reported by less than one-quarter of the cohort. When study did occur, the typical session duration was 1-2 hours (n = 72, 48.6%) or less than one hour (n = 57, 38.5%), with sessions lasting longer than two hours reported by only 12.8% of participants. Despite infrequent and brief study patterns, almost all respondents (n = 143, 96.6%) reported that they believed they needed to study more. The most frequently cited barrier was fatigue and burnout (n = 67, 46.9%), followed by time constraints due to clinical duties (n = 42, 29.4%), lack of motivation (n = 20, 14.0%), and personal responsibilities outside work (n = 14, 9.8%) (Table 1).

Regular use of AI tools for study purposes (2-3 times per week or more) was reported by 24.3% of participants (Table 2). Among those who did use AI tools, ChatGPT or similar LLMs were by far the most commonly accessed (n = 86, 58.1%), while the use of clinical decision-support systems, AI-powered medical databases, or medical image analysis tools was uncommon (≤ 8.1%). The most frequent purposes for AI use were information retrieval (n = 84, 56.8%), academic writing or presentation preparation (n = 41, 27.7%), and support with practice questions or case studies (n = 27, 18.2%).

The participants' perceptions of AI tools were characterized by moderate reliability [mean 3.22 ± 0.97, median 3.0 [(3.0-4.0)], higher perceived usefulness [mean 3.89 ± 0.99, median 4.0 (3.0-5.0)], and perceived time-saving efficiency [mean 4.01 ± 1.01, median 4.0 (3.0-5.0)]. Despite this, 67.6% of respondents rated AI tools as useful or very useful (scores of 4 or 5), yet actual usage remained limited, with 37.8% reporting no use at all (Table 2).

Textbooks were the most frequently reported learning resource (n = 137, 92.6%), followed by online articles or journals (n = 78, 52.7%) and educational videos (n = 63, 42.6%) (Table 3).

Among the 130 participants who had completed at least one ÇADS examination, the most frequently reported score category was 60-79 (95/130, 73.1%), while 26 participants (20.0%) reported scores between 80 and 100. Almost all participants (n = 142, 95.9%) had not yet attempted the EDAIC examinations (Table 3).

Exploratory analyses examined the relationship between study-related factors and ÇADS performance among the 130 participants

Table 1. Demographic and general characteristics of participants (n = 148).

Variable	n	%
Age group		
25-30	111	75.0
31-35	32	21.6
36-40	5	3.4
Sex		
Female	84	56.8
Male	64	43.2
Year of residency		
1st year	26	17.6
2nd year	55	37.2
3rd year	22	14.9
4th year	22	14.9
5th year	23	15.5
Type of institution		
City hospital	93	62.8
University hospital	37	25.0
Training & research hospital	18	12.2
Study frequency (outside clinical hours)		
Every day	5	3.4
2-3 times/week	28	18.9
Once a week	26	17.6
Occasionally	89	60.1
Average study session duration		
<1 hour	57	38.5
1-2 hours	72	48.6
2-3 hours	16	10.8
>3 hours	3	2.0
Perceives need for more studying		
Yes	143	96.6
No	5	3.4
Main barrier to increased study (n = 143)		
Fatigue and burnout	67	46.9
Time constraints due to clinical duties	42	29.4
Lack of motivation	20	14.0
Personal responsibilities outside work	14	9.8

Data are presented as number (n) and percentage (%). Year of residency reflects the self-reported training year at the time of the survey.

Table 2. Artificial intelligence tool use among anesthesiology trainees (n = 148).

Variable	n	%
Frequency of AI tool use for studying		
Every day	14	9.5
2-3 times/week	22	14.9
Once a week	8	5.4
Occasionally	48	32.4
Never	56	37.8
AI tools used (multiple response)		
ChatGPT or similar conversational AI	86	58.1
Clinical decision-support systems	1	0.7
AI-powered medical databases	12	8.1
Presentation-generating AI tools	6	4.1
Medical image analysis tools	8	5.4
Purposes of AI use (multiple response)		
Information retrieval	84	56.8
Practice questions/case studies	27	18.2
Academic writing/presentation	41	27.7
Exam preparation (EDAIC/ÇADS)	12	8.1
Simulation/guideline learning	22	14.9
Visual materials	25	16.9
AI evaluation scale (1-5)	**Mean ± SD**	**Median [IQR]**
Perceived reliability	3.22 (± 0.97)	3.0 [3.0-4.0]
Perceived usefulness	3.89 (± 0.99)	4.0 [3.0-5.0]
Time-saving efficiency	4.01 (± 1.01)	4.0 [3.0-5.0]

Data for frequency and tools are presented as n (%) or mean ± SD and median [IQR] as appropriate. Multiple-response items sum to more than 100%. AI evaluation scales: 1 = strongly disagree, 5 = strongly agree.
SD: Standard deviation, IQR: Interquartile range, AI: Artificial intelligence, EDAIC: European Diploma in Anaesthesiology and Intensive Care, ÇADS: *Çevrimiçi Asistan Değerlendirme Sınavı*.

who had completed at least one ÇADS examination (Table 4). ÇADS score categories were treated as ordinal ranks (1-4) for statistical comparison. AI tool use frequency was significantly associated with higher ÇADS scores (Kruskal-Wallis H = 15.76, p = 0.003). Post-hoc pairwise comparisons revealed that daily users [median score 3 (IQR: 3-4); 33.3% scoring ≥ 80] and users reporting 2-3 times per week (median 3 [IQR: 3-4]; 42.1% scoring ≥ 80) performed significantly better than non-users (median 3 [IQR: 3-3]; 8.3% scoring ≥ 80) (Dunn's test: p = 0.006 and p = 0.013, respectively).

Self-reported confidence in clinical work was also positively associated with ÇADS performance (Spearman r_s = 0.205, p = 0.019). Trainees with very high self-confidence (score 5) had a median ÇADS score of 3.5 [IQR: 3-4], with 50.0% achieving scores ≥ 80, whereas those with very low confidence (score 1) had a median score of 2 [IQR 2-3], with none scoring ≥ 80.

No significant associations were found between ÇADS performance and study frequency outside clinical hours (p = 0.644) or the use of specific study techniques such as spaced repetition, note-taking, or case-based exercises (all p > 0.19).

To assess whether the observed associations between AI use and ÇADS performance might be confounded by other study-related variables, we examined the correlations between the key predictors. No significant correlation was found between AI use frequency and study frequency outside clinical hours (r_s = 0.067, p = 0.421), between AI use and self-confidence level (r_s = 0.124, p = 0.138), or between

Table 3. Study habits, examination participation, and score distribution (n = 148).

Variable	n	%
ÇADS participation		
Regular attendance	98	66.2
Occasional	32	21.6
Never	18	12.2
Most recent ÇADS score		
80-100	26	20.0
60-79	95	73.1
40-59	7	5.4
<40	2	1.5
Participants who had not taken ÇADS (n = 18) were excluded from this distribution		
EDAIC participation		
Have not sat yet	142	95.9
Sat but did not pass at least one part	4	2.7
Sat and passed	2	1.4
Preferred study resources (multiple response)		
Textbooks	137	92.6
Practice question banks	12	8.1
Educational videos	63	42.6
Online articles/journals	78	52.7
AI tools	47	31.8
Group study	4	2.7
Study techniques used (multiple response)		
Question solving	13	8.8
Spaced repetition	90	60.8
Note-taking	81	54.7
Case-based exercises	48	32.4
Flashcards/post-it notes	18	12.2
Group study	7	4.7

AI: Artificial intelligence, ÇADS: *Çalışan Asistan Değerlendirme Sınavı* (trainee assessment examination conducted twice yearly by TARD), EDAIC: European Diploma in Anaesthesiology and Intensive Care. Multiple-response items sum to more than 100%.
Percentages for ÇADS score categories were calculated among participants who had completed at least one ÇADS examination (n = 130).

self-confidence and study frequency (r_s = 0.089, p = 0.285). These exploratory analyses showed that both frequent AI use and higher self-confidence were associated with higher ÇADS score categories. However, because no multivariable adjustment was performed, these findings should be interpreted as exploratory associations rather than evidence of independent effects and require confirmation in future prospective studies using adjusted statistical models.

DISCUSSION

This national survey of anesthesiology trainees in Türkiye revealed three key findings: first, that self-reported study outside clinical hours is infrequent and brief, with the majority of respondents citing fatigue and workload as primary barriers; second, that AI tools are perceived as useful but remain underutilized, with only one in four trainees reporting regular use; and third, that frequent AI use and higher self-confidence were associated with higher ÇADS score categories in exploratory univariable analyses, whereas traditional study habits showed no such association.

The finding that 60% of respondents study only "occasionally" and that typical study sessions last one to two hours or less aligns with previous reports describing limited and irregular self-directed learning among medical trainees under high clinical workloads (15). In a survey of radiology trainees in the United States, Toy et al. (1) similarly reported that most trainees studied less than two hours per day, with time constraints cited as the primary limiting factor. Likewise, studies from Japan and other countries have documented inverse associations between clinical duty hours and available study time, suggesting that service demands directly compete with educational activities (5,16).

What distinguishes our cohort is the prominence of fatigue and burnout over time constraints alone. Nearly half of the respondents (46.9%) identified fatigue and burnout as the main barriers to increased study, a higher proportion than those citing time limitations due to clinical duties (29.4%). This pattern may reflect the intense on-call structure and unpredictable workload characteristics of anesthesiology training in Türkiye (15). Previous work has shown that burnout is associated not only with reduced study time but also with lower motivation, impaired concentration, and decreased long-term retention (16). Notably, this pattern differs from recent findings in countries with regulated duty hours. Following the implementation of work-hour restrictions in the United States, personal and family responsibilities, rather than fatigue, emerged as the primary barrier to study time among trainees (17). Nearly all respondents (96.6%) reported needing to study more, despite minimal actual study, suggesting a cycle in which workload and exhaustion reduce both the capacity and the effectiveness of self-directed learning.

Although two-thirds of the respondents in the present study rated AI tools as useful or very useful, only 24% reported using them regularly (2-3 times per week or more) for study purposes, and 38% reported never using them at all. This gap between perceived utility and actual use is consistent with early adoption patterns observed in other educational technologies.

Several factors may contribute to this discrepancy. First, moderate reliability scores (mean 3.2 out of 5) suggest that trainees remain cautious about the accuracy of AI-generated content. Concerns about factual errors, outdated information, or inappropriate clinical recommendations in LLM outputs have been widely documented (10,11,18,19). Although tools such as ChatGPT have demonstrated passing-level performance on components of medical licensing examinations, performance on structured test questions does not guarantee reliability in open-ended clinical reasoning or nuanced case interpretation (11,12). Trainees may be understandably hesitant to rely on AI tools for high-stakes learning without external validation.

Second, a lack of familiarity or institutional guidance may limit uptake. Unlike traditional textbooks or peer-reviewed journals, AI tools, such as ChatGPT, do not come with established protocols

Table 4. Factors associated with ÇADS score: Comparative analysis (n = 130).

Variable/comparison	n	Median [IQR] ÇADS score[†]	% scoring 80-100	p-value[a]
A. Frequency of AI tool use (Kruskal-Wallis)				
Daily*	12	3 [3-4]	33.3	**0.0034**
2-3 times/week**	19	3 [3-4]	42.1	
Once a week	8	3.5 [3-4]	50.0	
Occasionally	43	3 [3-3]	14.0	
Never	48	3 [3-3]	8.3	
B. Study frequency outside clinical hours (Kruskal-Wallis)				
Every day	5	3 [3-3]	20.0	0.6443
2-3 times/week	25	3 [3-3]	28.0	
Once a week	22	3 [3-3]	13.6	
Occasionally	78	3 [3-3]	19.2	
C. Study technique use (Mann-Whitney U per technique)				
Technique	**Used n**	**Median (users vs. non-users)**	**% scoring 80-100 (users)**	**p-value[b]**
Question solving	13	3.0 vs. 3.0	30.8	0.1900
Spaced repetition	78	3.0 vs. 3.0	17.9	0.4240
Case-based exercises	42	3.0 vs. 3.0	26.2	0.1959
Note-taking	73	3.0 vs. 3.0	20.5	0.6544
Flashcards	17	3.0 vs. 3.0	29.4	0.6498
Group study	7	3.0 vs. 3.0	42.9	0.3801
D. Self-confidence level × ÇADS score (Spearman correlation)				
Self-confidence level	**n**	**Median ÇADS score**	**% scoring 80-100**	**p-value[c]**
Very low (1)	9	2 [2-3]	0.0	**0.0194**
Low (2)	14	3 [3-3]	21.4	
Moderate (3)	65	3 [3-3]	13.8	
High (4)	34	3 [3-4]	29.4	
Very high (5)	8	3.5 [3-4]	50.0	

[†]: ÇADS scores encoded as ordinal ranks: 1 = < 40, 2 = 40-59, 3 = 60-79, 4 = 80-100. Median [IQR] scores represent the central rank and IQR of each group. Bold p-values indicate statistical significance ($p < 0.05$).
[a]: Kruskal-Wallis test, [b]: Mann-Whitney U test, [c]: Spearman rank correlation coefficient.
Post-hoc comparison of AI use frequency and ÇADS score (Dunn's test):
- Daily vs. never: p = 0.006*,
- 2-3/week vs. never: p = 0.013**.
No significant association was found between study technique, study frequency, institution type, or perceived barrier and ÇADS performance.
IQR: Interquartile range, AI: Artificial intelligence, ÇADS: *Çevrimiçi Asistan Değerlendirme Sınavı.*

for use in medical training. Trainees may be uncertain about how to integrate these tools effectively into their study routines or may lack awareness of "writing effective prompts," fact-checking, and source verification. Educational programs that provide training on the responsible use of AI tools—emphasizing both their strengths and their limitations—may help bridge this gap (20).

Exploratory univariable analyses demonstrated an association between frequent AI use and higher ÇADS score categories. Trainees who used AI tools daily or several times per week were significantly more likely to achieve scores in the top performance category (≥ 80) compared to non-users (33-42% vs. 8%, $p < 0.01$). Although AI use frequency was not significantly correlated with study frequency or self-confidence, these exploratory analyses do not establish an independent effect and should not be interpreted causally.

Several mechanisms may explain this finding. First, AI tools such as ChatGPT allow for rapid, on-demand retrieval of explanations, definitions, and summaries, which may facilitate just-in-time learning during clinical rotations when formal study time is scarce (21,22). Unlike textbooks, which require sustained reading, or online searches, which yield variable-quality results, conversational AI can provide immediate, synthesized responses tailored to specific queries. This may be particularly valuable in anesthesiology, where trainees encounter diverse clinical scenarios requiring the rapid integration of pharmacology, physiology, and procedural knowledge.

Second, interactive questioning through AI platforms may support active learning and self-testing, both of which are associated with improved retention compared to passive reading (23).

Third, it is possible that frequent AI users represent a subset of trainees who are generally more engaged in self-directed learning, more comfortable with digital tools, or more motivated to seek out supplementary resources. However, the lack of correlation between AI use and traditional study frequency (r_s = 0.07, p = 0.42) makes a simple selection effect less likely. Instead, AI use appears to function as a distinct learning behavior independent of conventional study habits.

However, this association should be interpreted cautiously. Because of the cross-sectional design, the direction of the relationship cannot be determined. Frequent AI use may contribute to learning; however, the reverse relationship is also plausible. Trainees with higher examination performance may be more motivated, more digitally literate, or more capable of using AI tools effectively and critically. Therefore, AI use should not be interpreted as a direct cause of better examination performance in the present study.

Nonetheless, the magnitude of the observed association suggests that the relationship warrants further investigation through prospective or interventional designs.

Self-reported confidence in clinical work showed a modest but statistically significant positive correlation with ÇADS performance (r_s = 0.21, p = 0.02). This finding is consistent with previous studies linking self-efficacy to academic achievement in medical training (4,24). Trainees with higher confidence may be more likely to engage in challenging study activities and perform better under examination conditions.

However, the direction of causality remains unclear. Confidence may drive better performance, or successful performance may reinforce confidence, creating a positive feedback loop. Alternatively, both may be influenced by underlying factors such as prior academic preparation, clinical exposure, or quality of mentorship. The lack of correlation between confidence and study frequency in our data (r_s = 0.09, p = 0.29) suggests that confidence does not simply reflect greater study effort but may instead relate to clinical experience, teaching quality, or individual psychological traits (24).

Study Limitations

Several limitations should be acknowledged. First, the cross-sectional design precludes causal inference. The observed association between AI use and examination performance may reflect reverse causation, unmeasured confounding, or both. Second, all data were self-reported and therefore subject to recall bias and social desirability bias. ÇADS performance was based on self-reported score categories rather than verified official examination records. This may have introduced recall bias, selective reporting, or social desirability bias, particularly if trainees with higher scores were more willing to report their performance. Trainees may have overestimated their study time, underreported barriers, or selectively recalled AI use patterns. Third, ÇADS scores were available for only 130 of the 148 respondents, and performance was self-reported rather than verified through official records. Fourth, the survey used a convenience sample recruited through professional communication networks, which may have resulted in selection bias toward trainees who were more engaged, digitally active, or interested in AI-related topics. Because the total number of eligible trainees who received or viewed the survey invitation was unknown, a formal response rate could not be calculated. Consequently, non-response bias cannot be excluded, and the findings may not be fully representative of all anesthesiology trainees in Türkiye. Fifth, we did not assess specific patterns of AI use—such as the types of prompts used, the depth of engagement, or whether trainees verified AI-generated responses—limiting our ability to explain why frequent use was associated with better performance. Sixth, the questionnaire was developed specifically for this study and, although it underwent internal review and pilot testing, it was not a formally validated instrument. Finally, generalizability to other specialties, countries, or training systems may be limited. Because no multivariable analysis was performed, potential confounding factors could not be adjusted for, and the observed associations should therefore be interpreted with caution.

Our findings suggest that anesthesiology training programs in Türkiye should address workload and burnout as primary obstacles to effective learning, rather than focusing solely on providing more educational content. Interventions might include protected study time, wellness initiatives, or restructured duty schedules that prioritize recovery and sustained attention (25).

Regarding AI tools, the gap between perceived usefulness and actual use indicates a need for structured guidance. Training programs could introduce formal sessions on how to use AI tools effectively and safely, emphasizing prompt design, fact-checking, and appropriate use cases.

Future research should employ prospective or interventional designs to determine whether encouraging AI use directly improves learning outcomes or whether the observed association is driven by learner characteristics. Studies should also explore which specific patterns of AI use are most beneficial, how to mitigate reliability concerns, and whether AI tools can be integrated into formal curricula without displacing evidence-based learning strategies such as spaced repetition and active retrieval practice.

CONCLUSION

In conclusion, anesthesiology trainees in Türkiye report irregular and brief study patterns, with fatigue and clinical workload cited as the primary barriers to increased self-directed learning. Although AI tools are perceived as useful, adoption remains limited, likely due to moderate trust and a lack of guidance. Frequent AI use and higher self-confidence were associated with better performance in national in-training examinations, suggesting that these factors may be associated with better examination performance under time-constrained training conditions. Educational interventions aimed at reducing burnout while providing structured guidance for the responsible use of AI tools may improve learning environments for anesthesiology trainees.

Ethics

Ethics Committee Approval: Ethical approval was obtained from the Gazi University Ethics Commission (approval no: 13.01.2025-E.1141829, date: 13.01.2025).

Informed Consent: Participation was voluntary and anonymous, and all respondents provided electronic informed consent before proceeding to the survey.

Acknowledgments

Beyza Büyükgebiz Yeşil participated in the ESAIC Mentorship Programme as a mentee (ID: ESAIC_MSP_2024_BB).

Footnotes

Authorship Contributions

Concept: Ç.Ö., Design: Ç.Ö., Data Collection or Processing: M.G.Ö., Analysis or Interpretation: B.B.Y., Literature Search: B.B.Y., Writing: Ç.Ö.

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

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

REFERENCES

1. Toy D, Escalon JG, Groner LK, Naeger DM. Preparation for the ABR core exam: resident study habits and the value of case conferences. J Am Coll Radiol. 2021; 18: 615-9.
2. Liu A, Reed S, Mahan JD, Wallihan R. Exploring pediatric resident attitudes and preferences for board exam preparation. Cureus. 2020; 12: e8022.
3. Velez DR. Modern didactic formats in surgery: a systematic review. Am Surg. 2023; 89: 1701-8.
4. Barragán-Hervella RG, Gaytán-Fernández S, Ramírez-Polanco A, Martínez-Asención JP, Castillo-Ramírez IO, Tapia-Cortes C, et al. Study habits and academic performance in residents of orthopedics and family medicine. Rev Med Hosp Gen Mex. 2023; 86: 4-11.
5. Nagasaki K, Nishizaki Y, Shinozaki T, Kobayashi H, Tokuda Y. Association between resident duty hours and self-study time among postgraduate medical residents in Japan. JAMA Netw Open. 2021; 4: e210782.
6. Rao P, Jordan K, Burrows H, Helms L, Latch R, Perin R, et al. Association between ın-training exam scores and clinical exposure during the COVID-19 pandemic. Acad Pediatr. 2024; 24: 3-7.
7. Wilf-Miron R, Kagan I, Saban M. Health behaviors of medical students decline towards residency: how could we maintain and enhance these behaviors throughout their training. Isr J Health Policy Res. 2021; 10: 13.
8. Subramani M, Jaleel I, Krishna Mohan S. Evaluating the performance of ChatGPT in medical physiology university examination of phase I MBBS. Adv Physiol Educ. 2023; 47: 270-1.
9. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - Reshaping medical education and clinical management. Pak J Med Sci. 2023; 39: 605-7.
10. Lee H. The rise of ChatGPT: Exploring its potential in medical education. Anat Sci Educ. 2024; 17: 926-31. Erratum in: Anat Sci Educ. 2024; 17: 1779.
11. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ. 2023; 9: e45312. Erratum in: JMIR Med Educ. 2024; 10: e57594.
12. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023; 2: e0000198.
13. Li R, Kumar A, Chen JH. How chatbots and large language model artificial intelligence systems will reshape modern medicine: fountain of creativity or Pandora's box? JAMA Intern Med. 2023; 183: 596-7.
14. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res. 2004; 6: e34. Erratum in: doi:10.2196/jmir.2042.
15. Yakar MN, Bıyıkoğlu AS, Erbay Z, Öztürk MC, Öner Ö, Küçük M, et al. The effect of improvements in monthly income, employment conditions, and working hours on quality of life, anxiety, and burnout among anesthesiology residents: a before-after study. J Clin Pract Res. 2024; 46: 325-32.
16. Fafard J, Snell L. Reading habits of house-staff: what, where and why. Med Teach. 1989; 11: 279-83.
17. Edson RS, Beckman TJ, West CP, Aronowitz PB, Badgett RG, Feldstein DA, et al. A multi-institutional survey of internal medicine residents' learning habits. Med Teach. 2010; 32: 773-5.
18. Çamlar M, Sevgi UT, Erol G, Karakaş F, Doğruel Y, Güngör A. Comparative performance of neurosurgery-specific, peer-reviewed versus general AI chatbots in bilingual board examinations: evaluating accuracy, consistency, and error minimization strategies. Acta Neurochir (Wien). 2025; 167: 241.
19. Waldock WJ, Zhang J, Guni A, Nabeel A, Darzi A, Ashrafian H. The accuracy and capability of artificial intelligence solutions in health care examinations and certificates: systematic review and meta-analysis. J Med Internet Res. 2024; 26: e56532.
20. Xu Y, Jiang Z, Ting DSW, Kow AWC, Bello F, Car J, et al. Medical education and physician training in the era of artificial intelligence. Singapore Med J. 2024; 65: 159-66.
21. Mondal H, Karri JKK, Ramasubramanian S, Mondal S, Juhi A, Gupta P. A qualitative survey on perception of medical students on the use of large language models for educational purposes. Adv Physiol Educ. 2025; 49: 27-36.
22. Banerjee M, Chiew D, Patel KT, Johns I, Chappell D, Linton N, et al. The impact of artificial intelligence on clinical education: perceptions of postgraduate trainee doctors in London (UK) and recommendations for trainers. BMC Med Educ. 2021; 21: 429.
23. Luo X, Alias BS, Adnan NH. Active learning strategies and student engagement in undergraduate medical education: a systematic review (2015-2025). Anat Sci Educ. 2026; 19: 1157-69.
24. Patenaude B, Hogan RE, Manguvo A. Exploring the dynamics of medical students' exam performance in relation to study habits, growth mindset, confidence levels, and demographics. Med Sci Educ. 2024; 34: 371-8.
25. Martinelli SM, Tran TN, Canter CL, Hu D, Isaak RS, Afroze F, et al. Educating anesthesiology residents and their support persons to improve understanding and ıncrease wellness: a multisite study. Anesth Analg. 2026.

Original Investigation | Özgün Araştırma

Gazi Med J 2026;37(4):504-510

DOI: http://dx.doi.org/10.12996/gmj.2026.4731

Reliability of an AI-based Conversational Agent in Pediatric Dentofacial Health Counseling

Pediatrik Dentofasiyal Sağlık Danışmanlığında Yapay Zeka Tabanlı Konuşma Ajanının Güvenilirliği

Ceren Yapar Gümüş[1], Süleyman Kutalmış Büyük[2,3], Hale Akdemir[3]

[1]Department of Pediatrics, Ordu University, Faculty of Medicine, Ordu, Türkiye
[2]Department of Dental Biomaterials Science, Dental Research Institute, Seoul National University, Seoul, South Korea
[3]Department of Orthodontics, Ordu University, Faculty of Dentistry, Ordu, Türkiye

ABSTRACT

Objective: To evaluate the reliability and quality of Grok artificial intelligence's (AI) responses to commonly asked questions about pediatric dentofacial health.

Methods: Four queries related to detrimental oral habits were submitted to Grok AI in separate sessions. The generated responses were independently evaluated by orthodontists, pediatricians, and parents without a medical or dental background. Reliability was assessed using the eight-item modified DISCERN instrument, and overall information quality was evaluated using the Global Quality Scale (GQS). Group comparisons were performed using the Kruskal-Wallis test, and correlations among scores were analyzed using Spearman's rank correlation.

Results: Across evaluator groups, mean modified DISCERN scores ranged from 23.85 to 27.66, and mean GQS scores ranged from 3.61 to 4.12, corresponding, overall, to moderate-to-good reliability and quality. No statistically significant differences were observed among orthodontists, pediatricians, and parents for any query. Correlation analyses showed significant positive associations among several modified DISCERN and GQS parameters, particularly across different queries within the same evaluator groups. These findings suggested consistent scoring patterns, especially among parents and pediatricians.

Conclusion: Grok AI provided moderately reliable and generally useful responses regarding pediatric dentofacial health. However, its responses should be considered as supportive educational information

ÖZ

Amaç: Bu çalışmanın amacı, Grok yapay zeka (YZ) sisteminin pediyatrik dentofasiyal sağlık konusunda sık sorulan sorulara verdiği yanıtların güvenilirliğini ve kalitesini değerlendirmektir.

Yöntemler: Zararlı oral alışkanlıklarla ilişkili dört farklı soru, ayrı oturumlarda Grok YZ sistemine yöneltildi. Elde edilen yanıtlar; ortodontistler, çocuk sağlığı ve hastalıkları uzmanları ile tıp veya diş hekimliği alanında eğitim almamış ebeveynler tarafından birbirinden bağımsız olarak değerlendirildi. Yanıtların güvenilirliği, sekiz maddeden oluşan modifiye DISCERN aracı kullanılarak; genel bilgi kalitesi ise Global Quality Scale (GQS) ile değerlendirildi. Gruplar arasındaki karşılaştırmalar Kruskal-Wallis testi ile yapıldı, puanlar arasındaki ilişkiler ise Spearman sıra korelasyon analizi kullanılarak incelendi.

Bulgular: Değerlendirici gruplar genelinde ortalama modifiye DISCERN puanları 23,85 ile 27,66 arasında, ortalama GQS puanları ise 3,61 ile 4,12 arasında değişti. Bu değerler, yanıtların genel olarak orta ile iyi düzey arasında güvenilirlik ve kaliteye sahip olduğunu gösterdi. Hiçbir soru için ortodontistler, çocuk sağlığı ve hastalıkları uzmanları ve ebeveynler arasında istatistiksel olarak anlamlı bir fark saptanmadı. Korelasyon analizlerinde, çeşitli modifiye DISCERN ve GQS parametreleri arasında, özellikle aynı değerlendirici gruplarının farklı sorulara verdikleri puanlar açısından anlamlı pozitif ilişkiler belirlendi. Bu bulgular, özellikle ebeveynler ve çocuk sağlığı ve hastalıkları uzmanları arasında puanlama örüntülerinin tutarlı olduğunu düşündürdü.

Sonuç: Grok YZ, pediyatrik dentofasiyal sağlıkla ilgili konularda orta düzeyde güvenilir ve genel olarak yararlı yanıtlar sundu. Bununla birlikte,

Cite this article as: Yapar Gümüş C, Kutalmış Büyük S, Akdemir H. Reliability of an AI-based conversational agent in pediatric dentofacial health counseling. Gazi Med J. 2026;37(4):504-510

Address for Correspondence/Yazışma Adresi: Ceren Yapar Gümüş, Department of Pediatrics, Ordu University, Faculty of Medicine, Ordu, Türkiye
E-mail / E-posta: cerenyapar91@gmail.com
ORCID ID: orcid.org/0000-0001-6349-2514

Received/Geliş Tarihi: 03.06.2026
Accepted/Kabul Tarihi: 13.08.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

rather than a substitute for a professional pediatric or orthodontic assessment. Further validation using larger question sets, multiple AI models, and more diverse evaluator groups is needed.

Keywords: Artificial intelligence, generative artificial intelligence, natural language processing, child health education, malocclusion

ÖZ

bu yanıtlar profesyonel pediatrik veya ortodontik değerlendirmenin yerine geçen bir kaynak olarak değil, destekleyici eğitim bilgisi olarak değerlendirilmelidir. Daha geniş soru setlerinin, farklı YZ modellerinin ve daha çeşitli değerlendirici gruplarının yer aldığı ileri doğrulama çalışmalarına ihtiyaç vardır.

Anahtar Sözcükler: Yapay zeka, üretken yapay zeka, doğal dil işleme, çocuk sağlığı eğitimi, maloklüzyon

INTRODUCTION

Artificial intelligence (AI) applications, particularly chatbots powered by large language models, are being increasingly adopted in healthcare. These systems offer patients round-the-clock accessibility, user-friendliness, and rapid responses, enabling individuals to obtain information at any time of day (1,2). Owing to these features, they are regarded as a practical source of information by both the public and healthcare professionals.

Nevertheless, the accuracy and reliability of AI-based chatbots remain a matter of debate. Several studies have demonstrated that such tools may occasionally generate incorrect or irrelevant content and may even present misinformation as factual (3,4). More recent investigations have further highlighted that outputs generated by AI models can contain biases related to gender, race, or socioeconomic status (5). In addition, chatbot responses are often produced at a level exceeding recommended health literacy standards, which may hinder patients' understanding and reduce their active participation in care (6-8).

Recent studies have begun to evaluate the reliability of AI chatbots across different medical disciplines. Kurt Demirsoy et al. (9) examined ChatGPT's responses in orthodontics and found that, while the model could contribute to patient education, it remained limited in terms of factual accuracy. Similarly, Terzi et al. (10) reported that ChatGPT's responses regarding dental implants and peri-implant phenotypes were accessible but lacked sufficient depth for clinical decision-making. These findings underscore the potential utility of AI-based chatbots and emphasize the need for systematic evaluation across diverse clinical contexts. Although Grok has recently been included in comparative evaluations of large language models in dentistry, its reliability in pediatric dentofacial health counseling remains insufficiently characterized (11). This gap provided the rationale for selecting Grok for the present study.

Dentofacial health is closely linked to growth and development during childhood and adolescence. Harmful oral habits can negatively influence jaw development, disrupt muscle balance, and impair orofacial functions (12). In this context, both parents and healthcare professionals—particularly pediatricians and orthodontists—play a critical role in guiding healthy development. Preventive strategies and effective counseling, however, depend on access to trustworthy information (13). The reliability of information provided by AI-based digital advisors is therefore of great importance, not only for empowering parents but also for supporting clinicians in their educational and counseling roles.

This study aimed to assess the reliability of Grok AI, an AI-based chatbot, in providing guidance on dentofacial health. The chatbot's responses were evaluated by participants from different professional backgrounds using the modified DISCERN (mDISCERN) tool and the Global Quality Scale (GQS). By exploring the reliability of AI-based conversational agents in this field, the study seeks to contribute to both clinical practice and patient and parent education.

MATERIALS AND METHODS

In this study, frequently asked questions about detrimental oral habits were submitted to Grok AI, an artificial- intelligence-based conversational agent. To minimize the risk that the system's adaptive learning would influence subsequent responses, each question was entered separately into new sessions. The generated answers were then evaluated by three distinct participant groups through an online survey. Each group reviewed the same set of responses, thereby allowing a comprehensive assessment of Grok AI's accuracy, clarity, and relevance to the topic.

The following questions were directed to the chatbot:

1. What are the effects of mouth breathing on the dentofacial region?
2. What are the impacts of prolonged pacifier and bottle use on the head and facial region?
3. How do habits such as nail-biting, thumb-sucking, and biting foreign objects (e.g., pencils) affect the dentofacial region?
4. What are the effects of frenulum and tongue-tie on the dentofacial region?

The responses generated by Grok AI were assessed by three groups, including two expert panels, to capture perspectives across different demographics and educational backgrounds. The participant groups were defined as follows:

1. Pediatricians (medical specialists in child health),
2. Orthodontists (dental specialists in orthodontics),
3. Parents: individuals with at least an undergraduate degree, who have children but no professional background in medicine or dentistry.

Inclusion Criteria

- Having completed specialty training in pediatrics or orthodontics and voluntarily consenting to participate in the survey.
- For inclusion in the parent group: having at least an undergraduate degree, being a parent, and not belonging to a health-related profession.

Exclusion Criteria

- Being a resident in pediatrics or orthodontics.
- Declining to provide informed consent for participation.
- Parents who, despite meeting the above criteria, were affiliated with medical or dental professions.

The sample size was calculated using the G*Power program (version 3.1.9.2; Axel Buchner, University of Düsseldorf, Düsseldorf, Germany) to detect a medium effect size with 90% power. The required sample size was determined to be 66 participants (effect size: 0.4382637).

Evaluation of Responses

The chatbot's answers were evaluated using two validated instruments: the mDISCERN questionnaire and the GQS. Assessments were performed independently by three groups: orthodontists, pediatricians, and parents with no professional background in healthcare.

Evaluation Scales

The DISCERN instrument is a widely used three-part tool designed to evaluate the reliability and quality of online health information (3,14). In this study, only the first section consisting of eight items was applied to examine the reliability of responses generated by the AI-based chatbot (14). The eight criteria assessed in the mDISCERN scale were as follows: (i) clarity of aims, (ii) achievement of stated aims, (iii) relevance of content to the topic, (iv) citation of information sources, (v) currency of information, (vi) impartiality and balance, (vii) provision of additional sources of support, and (viii) acknowledgment of areas of uncertainty. Each item was rated on a scale from 1 (no) to 5 (yes). The total score was categorized as poor (8-15), moderate (16-31), and good (32-40) (15).

The GQS is a five-point instrument used to assess the quality of medical information sources (2). The scale is defined as follows:

- 1 point: poor quality, lacking information, not useful for patients,
- 2 points: poor quality, many important details missing, limited usefulness,
- 3 points: moderate quality, some important information adequately addressed, others insufficiently covered,
- 4 points: good quality, most key information adequate and useful for patients,
- 5 points: excellent quality, fluent, comprehensive, and highly useful.

Accordingly, GQS scores of 1-2 indicate low quality, scores of 3 indicate moderate quality, and scores of 4-5 indicate high quality.

This study was approved by the Ordu University Non-Interventional Scientific Research Ethics Committee (approval no: 2025/246, date: 18.07.2025). Electronic informed consent was obtained from all participants before participation in the online survey.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, version 20.0 (IBM Corp., Armonk, NY, USA). The normality of mDISCERN and GQS score distributions was assessed using the Shapiro-Wilk test. Descriptive statistics were presented as mean ± standard deviation and minimum-maximum values. Because the score distributions did not meet normality assumptions and the GQS represents an ordinal rating scale, nonparametric methods were used. The Kruskal-Wallis test was applied to compare mDISCERN and GQS scores among the three evaluator groups. Spearman's rank correlation analysis was used to evaluate associations between mDISCERN and GQS scores within each evaluator group. All statistical tests were two-tailed, and a p-value of < 0.05 was considered statistically significant.

RESULTS

The quality and reliability of Grok AI's responses were evaluated using the mDISCERN and GQS instruments across three participant groups: orthodontists, pediatricians, and parents. As shown in Table 1, the mean mDISCERN scores for the four queries did not differ significantly among the groups. The mean mDISCERN scores ranged from 23.85 to 24.90 in the orthodontist group, from 25.95 to 27.66 in the pediatrician group, and from 25.52 to 27.09 in the parent group. Similarly, GQS scores showed no statistically significant differences among the three groups, with mean values ranging from 3.61 to 4.12. These findings indicate that Grok AI's responses were rated at comparable levels by orthodontists, pediatricians, and parents, corresponding overall to moderate-to-good levels of reliability and quality.

Table 1. Comparison of Grok AI responses by mDISCERN and GQS scores across groups.

	Parameters	Orthodontists (n = 21)		Pediatricians (n = 24)		Parents (n = 21)		
		Mean (SD)	Min-max	Mean (SD)	Min-max	Mean (SD)	Min-max	p*
Query 1	mDISCERN	23.85 (3.38)	18-32	25.95 (4.17)	21-39	26.19 (5.06)	18-40	0.120
	GQS	3.71 (0.64)	2-5	4.04 (0.62)	3-5	4.09 (0.70)	3-5	0.166
Query 2	mDISCERN	24.90 (3.68)	20-37	27.66 (5.97)	19-40	27.09 (5.44)	18-40	0.175
	GQS	3.85 (0.72)	2-5	4.04 (0.75)	3-5	4.09 (0.83)	3-5	0.626
Query 3	mDISCERN	24.76 (3.78)	15-32	26.54 (5.83)	18-40	26.66 (5.62)	18-40	0.588
	GQS	3.71 (0.84)	2-5	3.95 (0.75)	3-5	4.04 (0.74)	3-5	0.470
Query 4	mDISCERN	24.85 (3.81)	18-36	27.33 (5.53)	18-40	25.52 (4.63)	18-40	0.248
	GQS	3.61 (0.80)	2-5	4.12 (0.67)	3-5	4.00 (0.77)	3-5	0.112

*: Results of Kruskal-Wallis test.
GQS: Global Quality Scale, SD: Standard deviation, Max: Maximum, Min: Minimum, mDISCERN: Modified DISCERN.

In the orthodontist group (Table 2), significant positive correlations were observed primarily among scores from the same assessment instrument across different items. Among GQS scores, the strongest correlations were found between Query 2 and Query 3 (r = 0.855, p < 0.001), Query 2 and Query 4 (r = 0.821, p < 0.001), and Query 3 and Query 4 (r = 0.792, p < 0.001). Similarly, mDISCERN scores showed significant positive correlations across different queries, including Query 2 and Query 3 (r = 0.756, p < 0.001), Query 2 and Query 4 (r = 0.747, p < 0.001), and Query 3 and Query 4 (r = 0.719, p < 0.001). These findings suggest that orthodontists exhibit a consistent scoring pattern when evaluating Grok AI's responses across different dentofacial health topics.

In the pediatrician group (Table 3), significant positive correlations were also observed among GQS scores across different queries. These correlations ranged from r = 0.616 to r = 0.865, with the strongest association observed between GQS scores for Query 3 and

Table 2. Correlation of GQS and mDISCERN scores in orthodontist (n = 21).

		Query 1 GQS	Query 1 mDISCERN	Query 2 GQS	Query 2 mDISCERN	Query 3 GQS	Query 3 mDISCERN	Query 4 GQS
Query 1 mDISCERN	r	0.053						
	p	0.820						
Query 2 GQS	r	**0.551***	0.081					
	p	0.010	0.727					
Query 2 mDISCERN	r	-0.289	**0.568***	0.091				
	p	0.205	0.007	0.696				
Query 3 GQS	r	0.414	0.107	**0.855***	0.142			
	p	0.062	0.646	< 0.001	0.540			
Query 3 mDISCERN	r	-0.077	**0.728***	0.034	**0.756***	0.141		
	p	0.739	< 0.001	0.882	< 0.001	0.542		
Query 4 GQS	r	**0.634***	0.091	**0.821***	0.103	**0.792***	-0.054	
	p	0.002	0.696	< 0.001	0.656	< 0.001	0.817	
Query 4 mDISCERN	r	0.064	**0.598***	0.330	**0.747***	0.383	**0.719***	0.372
	p	0.782	0.004	0.144	< 0.001	0.087	< 0.001	0.097

Bold values are statistically significant. *: Correlation is significant at the 0.01 level (2-tailed).
GQS: Global Quality Scale, mDISCERN: Modified DISCERN.

Table 3. Correlation of GQS and mDISCERN scores in pediatricians (n = 24).

		Query 1 GQS	Query 1 mDISCERN	Query 2 GQS	Query 2 mDISCERN	Query 3 GQS	Query 3 mDISCERN	Query 4 GQS
Query 1 mDISCERN	r	0.122						
	p	0.571						
Query 2 GQS	r	**0.657****	0.394					
	p	< 0.001	0.057					
Query 2 mDISCERN	r	0.165	**0.835****	**0.588****				
	p	0.441	< 0.001	0.003				
Query 3 GQS	r	**0.653****	0.264	**0.698****	**0.417***			
	p	0.001	0.213	< 0.001	0.043			
Query 3 mDISCERN	r	0.297	**0.797****	**0.521****	**0.792****	**0.529****		
	p	0.158	< 0.001	0.009	< 0.001	0.008		
Query 4 GQS	r	**0.629****	0.202	**0.616****	0.399	**0.865****	**0.415***	
	p	0.001	0.343	0.001	0.053	< 0.001	0.044	
Query 4 mDISCERN	r	0.149	**0.862****	0.446*	**0.867****	0.373	**0.843****	0.357
	p	0.488	< 0.001	0.029	< 0.001	0.073	< 0.001	0.087

*: Correlation is significant at the 0.05 level (2-tailed). Bold values are statistically significant. **: Correlation is significant at the 0.01 level (2-tailed).
GQS: Global Quality Scale, mDISCERN: Modified DISCERN.

Query 4 (r = 0.865, p < 0.001). Strong positive correlations were also detected among mDISCERN scores across different queries, ranging from r = 0.792 to r = 0.867. In addition, significant correlations were found between GQS and mDISCERN scores for Query 2 (r = 0.588, p = 0.003) and for Query 3 (r = 0.529, p = 0.008). These findings indicate that pediatricians demonstrate a consistent evaluation pattern, particularly in terms of response quality and reliability across the four queries.

Among parents (Table 4), GQS scores across different queries showed very strong positive correlations. The strongest correlations were detected between Query 2 and Query 4 (r = 0.928, p < 0.001), Query 2 and Query 3 (r = 0.884, p < 0.001), and Query 3 and Query 4 (r = 0.871, p < 0.001). Similarly, mDISCERN scores showed strong positive correlations across different queries, including Query 2 and Query 3 (r = 0.892, p < 0.001); Query 3 and Query 4 (r = 0.868, p < 0.001); Query 1 and Query 4 (r = 0.852, p < 0.001); Query 1 and Query 3 (r = 0.831, p < 0.001); Query 2 and Query 4 (r = 0.780, p < 0.001); and Query 1 and Query 2 (r = 0.677, p = 0.001). A significant positive correlation was also found between Query 3 GQS and Query 3 mDISCERN scores (r = 0.447, p = 0.042). These results suggest that parents tended to evaluate the quality and reliability of Grok AI's responses consistently across different questions.

Grok AI's responses were rated similarly by all three evaluator groups, with no significant between-group differences in mDISCERN or GQS scores. Correlation analyses revealed significant positive associations among several GQS and mDISCERN parameters, particularly across different queries evaluated by the same evaluator groups. These findings suggest consistency in evaluators' scoring patterns, particularly among parents and pediatricians.

DISCUSSION

In this study, the reliability and quality of responses generated by Grok AI, an AI-based conversational agent, were evaluated in the context of pediatric dentofacial health, based on assessments by orthodontists, pediatricians, and parents. The findings demonstrated that both mDISCERN and GQS scores exhibited overall moderate-to-good reliability and quality. Mean mDISCERN scores ranged between 23 and 27, while GQS scores ranged from 3.6 to 4.1, with no statistically significant differences observed among the three groups. These results suggest that Grok AI's responses were perceived as broadly comparable by both professional and non-professional evaluators.

Correlation analyses provided additional insight into evaluator scoring patterns. Significant positive correlations were observed among several GQS and mDISCERN parameters, particularly across different queries evaluated by the same evaluator groups. In the orthodontist group, correlations were primarily observed within GQS scores and within mDISCERN scores across different queries. In the pediatrician group, within-scale correlations and some GQS–mDISCERN correlations were significant. Among parents, very strong positive correlations were observed across GQS scores and mDISCERN scores for different queries. These findings suggest that participants consistently evaluated Grok AI's responses across dentofacial health topics. However, these results should be interpreted as evidence of consistent scoring patterns rather than complete agreement between the two instruments.

Similar findings have been reported in previous studies. Kurt Demirsoy et al. (9) examined ChatGPT's responses in the field of orthodontics and reported that, although the content was accessible, limitations remained in terms of factual accuracy. Similarly, Terzi et al. (10) evaluated AI-generated responses regarding dental implants and peri-implant phenotypes and concluded that, while the information may be useful for patient education, it lacked sufficient depth to support clinical decision-making. In line with these reports, our findings suggest that Grok AI may serve as a supportive educational resource to provide general information for parents. However, it

Table 4. Correlation of GQS and mDISCERN scores in parents (n = 21).

		Query 1 GQS	Query 1 mDISCERN	Query 2 GQS	Query 2 mDISCERN	Query 3 GQS	Query 3 mDISCERN	Query 4 GQS
Query 1 mDISCERN	r	0.331						
	p	0.142						
Query 2 GQS	r	**0.747****	0.405					
	p	< 0.001	0.069					
Query 2 mDISCERN	r	0.211	**0.677****	0.113				
	p	0.358	0.001	0.627				
Query 3 GQS	r	**0.750****	0.424	**0.884****	0.255			
	p	< 0.001	0.055	< 0.001	0.265			
Query 3 mDISCERN	r	0.308	**0.831****	0.347	**0.892****	**0.447***		
	p	0.174	< 0.001	0.124	< 0.001	0.042		
Query 4 GQS	r	**0.829****	0.325	**0.928****	0.078	**0.871****	0.263	
	p	< 0.001	0.151	< 0.001	0.735	< 0.001	0.249	
Query 4 mDISCERN	r	0.259	**0.852****	0.199	**0.780****	0.342	**0.868****	0.189
	p	0.256	< 0.001	0.386	< 0.001	0.129	< 0.001	0.412

*: Correlation is significant at the 0.05 level (2-tailed). Bold values are statistically significant. **: Correlation is significant at the 0.01 level (2-tailed).
GQS: Global Quality Scale, mDISCERN: Modified DISCERN.

should not be considered a substitute for professional pediatric or orthodontic assessment.

The general medical literature also supports similar conclusions. Liu et al. (16) demonstrated that, although ChatGPT offers practical benefits in generating medical content, issues of accuracy and currency remain. Giannakopoulos et al. (17) reported that large language models provide only limited reliability in the context of evidence-based dentistry. While Hwang et al. (18) highlighted the contribution of deep learning applications in dentistry, Yu et al. (19) emphasized that, despite the growing role of AI in clinical decision support systems, concerns regarding reliability and bias persist. Likewise, Topol (20) highlighted the transformative potential of AI in healthcare delivery but underscored the need to address ethical and transparency-related challenges. Taken together with our findings in pediatric dentofacial health, these observations support a cautious interpretation of AI-based conversational agents as adjunctive educational tools rather than as independent clinical decision-making resources.

One of the factors to consider when interpreting information generated by large language models is "hallucination." These models can occasionally generate incorrect or unsubstantiated information in a convincing manner; this can pose a problem, particularly when answering questions related to medicine or dentistry (21,22). Users may not always recognize such errors, especially when the provided response appears clear and authoritative. Although our study did not specifically evaluate hallucinations, the variability in reliability scores indicates that health information generated by AI should be treated with caution and, when necessary, verified by healthcare professionals.

Pediatrics is an important field for the evaluation of digital advisors, particularly because parental education and child health literacy may influence healthcare-seeking behavior and decision-making (23). Kubb and Foran (23), as well as Walsh et al. (24), demonstrated that parents frequently use online health information, which can significantly influence their decision-making processes. Baumann et al. (25) reported that parents seek child health information through multiple sources, including digital media, print media, and personal contacts, and that this behavior may vary according to whether the concern involves general child health and development or acute illness. Similarly, Min et al. (26), in a pediatric medical chatbot project developed in Singapore, emphasized that such tools may provide accurate, empathetic, and accessible information in urgent care contexts, although improvements are needed to better accommodate health literacy. DeWitt et al. (27) reviewed mobile health applications and noted that, although content was generally user-friendly, enhancements in readability and structure were required. Likewise, Klapow et al. (28) found that chatbot initiatives targeting parents were valuable for educational purposes but suggested that improvements in literacy adaptation and accuracy could further enhance user experience. Collectively, these findings underscore the growing role of digital tools in pediatric counseling. By directly including parents as evaluators, this study adds a parental perspective to assessing AI-generated pediatric dentofacial health information, although further research is needed to confirm and expand these findings.

Study Limitations

This study has several limitations. First, because the analysis was based on a limited number of queries, the findings may not reflect Grok AI's performance across the full spectrum of pediatric dentofacial health topics. Second, only one AI-based conversational agent was evaluated, and the lack of transparency regarding Grok AI's training data and update frequency may have influenced the accuracy and reproducibility of its responses. Third, the evaluations were subjective, and participants' varying levels of knowledge and experience may have affected the scoring. Moreover, inter-rater reliability was not assessed, which may limit interpretation of consistency across evaluators. The single-center design of the study may also limit the generalizability of the findings. Although the mDISCERN and GQS instruments are commonly used to assess the reliability and quality of online health information, both remain partially dependent on evaluators' interpretation. As the study was cross-sectional and based on hypothetical queries, it may not fully capture real-life interactions between parents, clinicians, and AI-based tools. In addition, each query was evaluated using a single generated response, and variations in Grok AI's responses across repeated sessions were not assessed. The evolving nature of large language models also means that responses may change over time, which could further limit reproducibility.

Future research should include comparative analyses of different AI models, larger and more diverse international samples, and direct assessments of patient experiences. In particular, the role of AI-based tools in parental education for pediatric populations warrants further investigation. Such efforts may clarify the potential contribution of AI-supported counseling to preventive care in addressing dentofacial health issues during childhood.

CONCLUSION

This study demonstrated that Grok AI provided responses with moderate-to-good reliability in the context of pediatric dentofacial health, as evaluated by orthodontists, pediatricians, and parents. These findings suggest that AI-based conversational agents may serve as supportive tools in parental education and preventive pediatrics. However, they remain insufficient for clinical decision-making, and further validation in larger and more diverse settings is warranted.

Ethics

Ethics Committee Approval: This study was approved by the Ordu University Non-Interventional Scientific Research Ethics Committee (approval no: 2025/246, date: 18.07.2025).

Informed Consent: Electronic informed consent was obtained from all participants before participation in the online survey.

Footnotes

Authorship Contributions

Concept: C.Y.G., S.K.B., Design: C.Y.G., S.K.B., Data Collection or Processing: C.Y.G., S.K.B., H.A., Analysis or Interpretation: C.Y.G., S.K.B., Literature Search: C.Y.G., S.K.B., H.A., Writing: C.Y.G., S.K.B., H.A.

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

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

REFERENCES

1. Abu Arqub S, Al-Moghrabi D, Allareddy V, Upadhyay M, Vaid N, Yadav S. Content analysis of AI-generated (ChatGPT) responses concerning orthodontic clear aligners. Angle Orthod. 2024; 94: 263-72.
2. Acar AH. Can natural language processing serve as a consultant in oral surgery? J Stomatol Oral Maxillofac Surg. 2024; 125: 101724.
3. Kılınç DD, Mansız D. Examination of the reliability and readability of Chatbot Generative Pretrained Transformer's (ChatGPT) responses to questions about orthodontics and the evolution of these responses in an updated version. Am J Orthod Dentofacial Orthop. 2024; 165: 546-55.
4. Eggmann F, Weiger R, Zitzmann NU, Blatz MB. Implications of large language models such as ChatGPT for dental medicine. J Esthet Restor Dent. 2023; 35: 1098-102.
5. Kim J, Cai ZR, Chen ML, Simard JF, Linos E. Assessing biases in medical decisions via clinician and AI chatbot responses to patient vignettes. JAMA Netw Open. 2023; 6: e2338050.
6. Pan A, Musheyev D, Bockelman D, Loeb S, Kabarriti AE. Assessment of artificial intelligence Chatbot responses to top searched queries about cancer. JAMA Oncol. 2023; 9: 1437-40.
7. Sriwastwa A, Ravi P, Emmert A, Chokshi S, Kondor S, Dhal K, et al. Generative AI for medical 3D printing: a comparison of ChatGPT outputs to reference standard education. 3D Print Med. 2023; 9: 21.
8. Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol. 2023; 29: 721-32.
9. Kurt Demirsoy K, Buyuk SK, Bicer T. How reliable is the artificial intelligence product large language model ChatGPT in orthodontics? Angle Orthod. 2024; 94: 602-7.
10. Terzi M, Yavuz MC, Bicer T, Buyuk SK. Evaluation of artificial intelligence robot's knowledge and reliability on dental implants and peri-implant phenotype. Sci Rep. 2025; 15: 9519.
11. Dave M, Tattar R, Alafaleg R, Barry S, Ariyaratnam S, Roudsari RV, et al. Performance of large language models (ChatGPT4-0, Grok2 and Gemini) in UK dentistry and dental hygiene and therapy assessments. Br Dent J. 2025.
12. Giannini L, Galbiati G, Cressoni P, Esposito L. Bad oral habits: a review of the literature. J Biol Regul Homeost Agents. 2021; 35: 403-6.
13. Hood CA, Hunter ML, Kingdon A. Demographic characteristics, oral health knowledge and practices of mothers of children aged 5 years and under referred for extraction of teeth under general anaesthesia. Int J Paediatr Dent. 1998; 8: 131-6.
14. Meade MJ, Dreyer CW. Web-based information on orthodontic clear aligners: a qualitative and readability assessment. Aust Dent J. 2020; 65: 225-32.
15. Onder CE, Koc G, Gokbulut P, Taskaldiran I, Kuskonmaz SM. Evaluation of the reliability and readability of ChatGPT-4 responses regarding hypothyroidism during pregnancy. Sci Rep. 2024; 14: 243.
16. Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. J Med Internet Res. 2023; 25: e48568.
17. Giannakopoulos K, Kavadella A, Aaqel Salim A, Stamatopoulos V, Kaklamanos EG. Evaluation of the performance of generative AI large language models ChatGPT, Google Bard, and Microsoft Bing Chat in supporting evidence-based dentistry: comparative mixed methods study. J Med Internet Res. 2023; 25: e51580.
18. Hwang JJ, Jung YH, Cho BH, Heo MS. An overview of deep learning in the field of dentistry. Imaging Sci Dent. 2019; 49: 1-7.
19. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018; 2: 719-31.
20. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019; 25: 44-56.
21. Gravel J, D'Amours-Gravel M, Osmanlliu E. Learning to fake it: limited responses and fabricated references provided by ChatGPT for medical questions. Mayo Clin Proc Digit Health. 2023; 1: 226-34.
22. Templin T, Perez MW, Sylvia S, Leek J, Sinnott-Armstrong N. Addressing 6 challenges in generative AI for digital health: a scoping review. PLOS Digit Health. 2024; 3: e0000503.
23. Kubb C, Foran HM. Online health information seeking by parents for their children: systematic review and agenda for further research. J Med Internet Res. 2020; 22: e19985.
24. Walsh AM, Hyde MK, Hamilton K, White KM. Predictive modelling: parents' decision making to use online child health information to increase their understanding and/or diagnose or treat their child's health. BMC Med Inform Decis Mak. 2012; 12: 144.
25. Baumann I, Jaks R, Robin D, Juvalta S, Dratva J. Parents' health information seeking behaviour - does the child's health status play a role? BMC Fam Pract. 2020; 21: 266.
26. Min C, Lim RXC, Tan SW, Ganapathy S. Experience developing a pediatric medical chatbot in Singapore: a digital innovation for improved emergency care. Front Digit Health. 2025; 7: 1557804.
27. DeWitt A, Kientz J, Liljenquist K. Quality of mobile apps for child development support: search in app stores and content analysis. JMIR Pediatr Parent. 2022; 5: e38793.
28. Klapow MC, Rosenblatt A, Lachman J, Gardner F. The feasibility and acceptability of using a digital conversational agent(chatbot)for delivering parenting interventions: systematic review. JMIR Pediatr Parent. 2024; 7: e55726.

Original Investigation | Özgün Araştırma

Gazi Med J 2026;37(4):511-517

DOI: http://dx.doi.org/10.12996/gmj.2026.4743

The Impact of Clinical Leadership Training on Internal Medicine Residents: A Design-based Quasi-experimental Study

Klinik Liderlik Eğitiminin İç Hastalıkları Asistanları Üzerindeki Etkisi: Tasarım Tabanlı Yarı Deneysel Bir Çalışma

Mehmet Koca[1,2], Özlem Coşkun[2]

[1]General Directorate of State Supply Office, Ankara, Türkiye
[2]Department of Medical Education, Gazi University, Faculty of Medicine, Ankara, Türkiye

ABSTRACT

Objective: This study aimed to assess the perceptions of clinical leadership among internal medicine trainees and to examine the impact of a structured clinical leadership training programme on learning outcomes.

Methods: The study was conducted between August 2023 and January 2024 in a tertiary hospital using a design-based research approach. Under this approach, which comprised needs analysis, programme development, implementation, and evaluation, 125 internal medicine residents were included: 63 participated in the 13-hour clinical leadership training programme (training group), while 62 formed the control group. Data were collected using a demographic form, a knowledge test (pre-test/post-test), and the clinical leadership scale (CLS). The chi-square test was used for between-group comparisons, and the odds ratio (OR) and 95% confidence interval (CI) were used to calculate the effect size. The Bonferroni correction was applied for multiple comparisons.

Results: The CLS demonstrated high internal consistency (Cronbach's alpha = 0.96). The highest scores were obtained in the "management of services" and "working with others" sub-dimensions, while the lowest score was recorded in the "designing interventions" sub-dimension. The correct response rate among participants who had received training was significantly higher than among those who had not (61.1% versus 29.2%), corresponding to a 2.09-fold increase (OR = 3.95; 95% CI: 3.22-4.85; $p < 0.001$).

ÖZ

Amaç: Bu çalışmada, iç hastalıkları asistanlarının klinik liderlik algılarının değerlendirilmesi ve yapılandırılmış bir klinik liderlik eğitim programının öğrenme çıktıları üzerindeki etkisinin incelenmesi amaçlanmıştır.

Yöntemler: Çalışma, Ağustos 2023-Ocak 2024 tarihleri arasında üçüncü basamak bir hastanede tasarım tabanlı araştırma yaklaşımıyla yürütülmüştür. İhtiyaç analizi, program geliştirme, uygulama ve değerlendirme aşamalarından oluşan bu yaklaşım kapsamında 125 iç hastalıkları asistanı çalışmaya dahil edilmiştir; bunlardan 63'ü 13 saatlik klinik liderlik eğitim programına katılmış (eğitim grubu), 62'si kontrol grubunu oluşturmuştur. Veriler demografik bilgi formu, bilgi testi (ön test/son test) ve klinik liderlik ölçeği (KLÖ) ile toplanmıştır. Gruplar arası karşılaştırmalarda ki-kare testi, etki büyüklüğünün hesaplanmasında olasılık oranı (OO) ve %95 güven aralığı (GA) kullanılmıştır. Çoklu karşılaştırmalar için Bonferroni düzeltmesi uygulanmıştır.

Bulgular: KLÖ'nün iç tutarlılığı yüksek bulunmuştur (Cronbach alfa = 0,96). En yüksek puanlar "hizmetlerin yönetimi" ve "diğerleri ile çalışma" alt boyutlarında, en düşük puan ise "yönlendirmelerin kurgulanması" alt boyutunda elde edilmiştir. Eğitim alan katılımcıların doğru yanıt oranı, eğitim almayanlara göre anlamlı düzeyde yüksek bulunmuştur (%61,1'e karşı %29,2); bu fark 2,09 katlık artışa karşılık gelmektedir (OO = 3,95; %95 GA: 3,22-4,85; $p < 0,001$).

Sonuç: Klinik liderlik eğitimi, iç hastalıkları asistanlarında bilgi ve farkındalık artışı ile ilişkilidir. Bu bulgular, klinik liderlik eğitim

Cite this article as: Koca M, Coşkun Ö. The impact of clinical leadership training on internal medicine residents: a design-based quasi-experimental study. Gazi Med J. 2026;37(4):511-517

Address for Correspondence/Yazışma Adresi: Mehmet Koca, General Directorate of State Supply Office; Department of Medical Education, Gazi University, Faculty of Medicine, Ankara, Türkiye
E-mail / E-posta: drmehmetkoca@gmail.com
ORCID ID: orcid.org/0000-0003-1108-3182

Received/Geliş Tarihi: 17.06.2026
Accepted/Kabul Tarihi: 16.07.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

Conclusion: Clinical leadership training is associated with increased knowledge and awareness among internal medicine residents. These findings support the integration of clinical leadership training programmes into the medical education curriculum.

Keywords: Clinical leadership, leadership training, graduate medical education, internal medicine residency, clinical competence, health professions education, quasi-experimental study

ÖZ

programlarının tıp eğitimi müfredatına entegre edilmesini desteklemektedir.

Anahtar Sözcükler: Klinik liderlik, liderlik eğitimi, mezuniyet sonrası tıp eğitimi, iç hastalıkları uzmanlık eğitimi, klinik yeterlilik, sağlık meslekleri eğitimi, yarı deneysel çalışma

INTRODUCTION

The increasing complexity of healthcare services and the rise in multidisciplinary teamwork have made the concept of clinical leadership a fundamental component of modern healthcare systems (1,2). Clinical leadership plays a critical role in improving the quality of patient care, ensuring coordination within teams, and supporting the effective management of healthcare services (3).

It has been demonstrated that effective clinical leadership is closely linked to patient safety, quality of care and healthcare performance (4). However, it is reported that clinical leadership skills are largely acquired through experience and that structured training programmes are limited (4,5).

Residents play an active role in patient care and assume significant responsibilities for team coordination and clinical decision-making. However, it is noted that current training programmes lack sufficient content for the systematic development of leadership skills (5,6). This shortcoming is of particular importance, as doctors are expected to manage clinical teams, implement quality improvements, and use resources effectively. Contrary to expectations, these skills are rarely taught during specialist training (7,8).

Although efforts to develop clinical leadership training programmes have increased in recent years, it is evident that there are limited studies objectively evaluating the impact of these programmes on learning outcomes (5). To the best of our knowledge, this study is the first to evaluate the effectiveness of a structured clinical leadership programme specifically designed for doctors in Türkiye. This study aimed to assess postgraduate trainees' perceptions of clinical leadership at University of Health Sciences Türkiye, Ankara Bilkent City Hospital and to examine the impact of a clinical leadership training programme for internal medicine postgraduate trainees on learning outcomes.

MATERIALS AND METHODS

Study Design and Setting

This study was conducted using a design-based research (DBR) approach to assess the clinical leadership perceptions of postgraduate trainees working at University of Health Sciences Türkiye, Ankara Bilkent City Hospital and to examine the impact of the clinical leadership training programme developed for internal medicine postgraduate trainees on learning outcomes (9). DBR is an iterative methodological framework comprising cycles of needs analysis, programme development, implementation, and evaluation. The research was conducted between August 2023 and January 2024.

During the needs-analysis phase of the study, an online survey was sent to approximately 1,000 postgraduate trainees at University of Health Sciences Türkiye, Ankara Bilkent City Hospital, and data were obtained from 442 participants. Among these, 63 internal medicine postgraduate trainees comprised the training group and participated in the clinical leadership training programme. The 62 internal medicine residents who did not participate in the training programme were selected from the same institution to form the control group (total n = 125). Both groups consisted of internal medicine residents with similar clinical responsibilities at the same hospital. Snowball sampling was used.

Training Programme

A 13-hour clinical leadership training programme comprising 12 topics was developed based on the findings of the needs analysis, a review of the literature, and expert opinions (Table 1). The first and final sessions of the programme were conducted face-to-face in the General Hospital Training Hall at University of Health Sciences Türkiye, Ankara Bilkent City Hospital, while the remaining sessions were delivered through an online learning system. For each session, the objectives, learning outcomes, content and duration were clearly defined. The training was delivered by trainers who are experts in their field.

Data Collection Tools

Three tools were used for data collection: (1) A demographic information form (age, gender, marital status, years of residency, and status regarding management, leadership, or clinical leadership training); (2) the clinical leadership scale (CLS); and (3) a knowledge test. The CLS consists of 40 items and five sub-dimensions, covering the following dimensions: Personal characteristics, working with others, management of services, improvement of services, and designing interventions (10).

Knowledge Assessment

Participants were administered a pre-test (14 questions) prior to the training and a post-test (13 questions) following the training. The same post-test was also administered to a control group that did not receive training, and comparisons were made between the groups. Although the pre-test and post-test are similar in content, they have different levels of difficulty, because they are not parallel forms.

Sample

The sample size for the study was determined using an a priori power analysis. For a clinically meaningful absolute difference of 30% in the proportion of correct answers on the knowledge test

between groups receiving and not receiving the clinical leadership intervention, the expected effect size was calculated as Cohen's h = 0.61, corresponding to a large effect. For calculations using G*Power 3.1 software to compare two independent proportions, a minimum of 42 participants per group was required under two-tailed $\alpha = 0.05$ and 80% statistical power. Allowing for a potential attrition rate of 30%, the target sample size was set at a minimum of 55 participants per group (11).

Statistical Analysis

Data were analysed using IBM SPSS Statistics 27 (IBM Corp., Armonk, NY, USA). Descriptive statistics are presented as frequency, percentage, mean and standard deviation. The reliability of the scale was assessed using Cronbach's alpha, and sample adequacy was evaluated using the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity. The correct response rates in groups with and without education were compared using the chi-square test. For each significant chi-square result, the odds ratio (OR) and 95% confidence interval (CI) were calculated as measures of effect size. The risk ratio was also calculated to compare the overall accuracy. To address multiple comparisons in the parallel chi-square tests for the 13 questions, a Bonferroni correction was applied, setting the corrected significance threshold at $p < 0.0038$ (0.05/13). The threshold for statistical significance was set at $p < 0.05$.

Ethics Committee

Ethical approval for the study was obtained from the University of Health Sciences Türkiye, Ankara Bilkent City Hospital Ethics Committee no. 2 via decision no: E2-22-2526, dated: 28 September 2022. Informed consent was obtained from all participants prior to participation. The scales used in the study were employed after obtaining the necessary permissions for their Turkish adaptations.

RESULTS

Descriptive Statistics

A total of 125 internal medicine trainees were included in the study. The demographic characteristics and training status of the participants are presented in Table 2.

The CLS demonstrated excellent internal consistency and sampling adequacy (Table 3). The overall Cronbach's alpha coefficient was 0.96, with subscale values ranging from 0.90 to 0.94. The KMO value was 0.906, and Bartlett's test of sphericity was statistically significant ($p < 0.001$), indicating that the data were suitable for factor analysis.

The subscale scores of the CLS are summarised in Table 4. The highest mean scores were observed in the "managing services" (83.94) and "working with others" (83.80) domains; the lowest mean score was observed in the 'designing interventions' domain (79.30).

Table 1. Clinical leadership training programme topic headings.

No	Topic	Presentation format
1	**What kind of leader are you? Definition of clinical leadership and types of leadership**	**Face-to-face**
2	The need for motivation and mentor support	Online
3	Emotional intelligence and working with different personalities	Online
4	The importance of a feedback culture in teamwork	Online
5	Time management	Online
6	Communication skills and a culture of empathy	Online
7	Anger management	Online
8	The art of public speaking in leadership	Online
9	Effective human resource management	Online
10	Challenges a clinical leader may face during healthcare delivery	Online
11	Pre- and post-graduate training practices in clinical leadership worldwide	Online
12	Crisis management and disaster management	Face-to-face (closing)

Table 2. Demographic characteristics and previous training status of internal medicine residents.

Variable	Category	n (%)
Gender	Female	76 (60.8)
	Male	49 (39.2)
Marital status	Single	60 (48.0)
	Married	65 (52.0)
Previous training	Management training (yes)	7 (5.6)
	Leadership training (yes)	5 (4.0)
	Clinical leadership training (yes)	3 (2.4)
Total		125 (100)

Intergroup Knowledge Test Performance

A comparison of correct answers on the final test between the trained and untrained groups is presented in Table 5. The trained group answered 500 out of 819 questions correctly (61.1%), while the untrained group answered 235 out of 806 questions correctly (29.2%). Overall, the performance of participants who received training was 2.09 times higher than that of those who did not receive training (OR = 3.95; 95% CI: 3.22-4.85; p < 0.001).

After applying the Bonferroni correction (threshold: p < 0.0038), the group that received training showed significantly better time-management performance (53 vs. 6; chi-square = 12.95; p = 0.0003; OR = 18.36; 95% CI: 7.50-44.94). However, the differences in motivation strategies (54 vs. 10; chi-square = 7.81; p = 0.0052) and in anger management (31 vs. 25; p = 0.0493) did not remain statistically significant. The group that did not receive training performed better on the question regarding challenges in healthcare leadership (39 vs.

Table 3. Reliability and sampling adequacy of the clinical leadership scale (KMO and Bartlett's test).

Scale/subscale	Number of items	Cronbach's alpha	
Clinical leadership scale (total)	40	0.96	
Personal characteristics subscale	8	0.91	
Working with others subscale	8	0.90	
Service management subscale	8	0.92	
Service improvement subscale	8	0.94	
Subscale for the construction of directions	8	0.93	
KMO value		0.906	
Bartlett's sphericity test			p < 0.001

KMO: Kaiser-Meyer-Olkin.

Table 4. Clinical leadership scale subscale scores of internal medicine residents.

Subscale	Number of items	Mean	Score (/100)	Standard deviation
Personal traits	8	2.47	82.28	0.58
Working with others	8	2.52	83.80	0.56
Managing services	8	2.52	83.94	0.56
Improving services	8	2.46	82.14	0.56
Setting direction	8	2.38	79.30	0.62
General scale	40	2.47	82.29	0.58

Table 5. Comparison of correct answers between trained and untrained internal medicine residents.

Question	Training group	Training not trained	Chi-square	p	OR (95% CI)
Motivation strategy	54	10	7.81	**0.0052***	9.61 (4.51-20.47)
Emotional intelligence in leadership	39	19	4.12	0.042	2.94 (1.48-5.84)
Leader's first step in a crisis	55	26	9.03	0.003	3.94 (1.95-7.96)
Leader's priority during a disaster	54	23	7.12	0.008	3.71 (1.83-7.52)
Importance of anger control	31	25	3.87	0.049	1.86 (0.93-3.72)
Improving communication skills	23	11	2.01	0.156	2.44 (1.08-5.49)
Impact of feedback culture	59	21	11.42	0.001	6.08 (2.70-13.67)
The role of rhetoric in leadership	17	8	1.59	0.207	2.40 (0.93-6.20)
Importance of merit-based human resources	33	25	0.89	0.345	1.59 (0.79-3.21)
Benefits of postgraduate training	20	5	4.78	0.029	4.67 (1.60-13.64)
Time management strategy	53	6	12.95	**0.0003****	18.36 (7.5-44.94)
Leadership style in a crisis	39	17	5.22	0.022	3.23 (1.55-6.74)
Challenges in healthcare leadership	23	39	28.25	<0.0001	0.12 (0.06-0.25)
Total	**500/819 (61.1%)**	235/806 (29.2%)	**166.8**	**< 0.001**	**OR = 3.95 (3.22-4.85)**

*: p < 0.05 (lost statistical significance after Bonferroni correction; did not meet the p < 0.0038 threshold), **: Remained statistically significant after Bonferroni correction (p < 0.0038), OR: Odds ratio, CI: Confidence interval

23; chi-square = 28.25; $p < 0.0001$; OR = 0.12; 95% CI: 0.06-0.25). No statistically significant differences were found between the groups on questions regarding emotional intelligence and crisis leadership (Bonferroni-corrected $p > 0.05$).

DISCUSSION

In this study, internal medicine trainees' perceptions of clinical leadership were generally high, with particularly high scores in the sub-dimensions of "management of services" and "working with others", and lower scores in the "designing interventions" sub-dimension. This finding suggests that trainees play an active role in team collaboration and service delivery within clinical practice, but have relatively limited experience in higher-level leadership skills, such as strategic thinking and setting direction. Significantly higher correct-response rates among participants who received clinical leadership training demonstrate the positive impact of structured training programmes on knowledge and awareness. The finding that clinical leadership training produces positive outcomes in junior doctors is consistent with the literature. Blumenthal et al. (7) comprehensively documented the gap in this area, emphasising the necessity of systematic leadership development programmes for junior doctors. Similarly, Sadowski et al. (5), in their systematic review of leadership interventions in postgraduate medical education, concluded that structured programmes enhance knowledge and awareness.

The first component of clinical leadership comprises personal qualities. In this study, the group that received training performed better in this dimension. This domain encompasses developing self-awareness, managing oneself, continuing personal development and acting with integrity (12). In their study on learning patterns, Santa and Nurcan (13) emphasised the impact of education on individual characteristics and noted that education provides awareness and insight into accepting environmental challenges and developing application patterns by altering psychological processes.

The second component of clinical leadership is the ability to work with others, i.e., suitability for teamwork. This study found a positive and significant relationship between the development of this dimension and education. Hunt et al. (14) found that teaching teamwork principles has a positive effect on enhancing the performance of team members, particularly in situations that require interdisciplinary communication. Harvey et al. (15) examined the effect of teamwork on nurses' performance and found that providing training in teamwork to nurses improved both their performance and their attitudes towards teamwork. Effective leadership requires people to work together to improve services. Consistent with this view, our study finds a positive relationship between the development of this competence and education. Clinical leaders must be effective in promoting communication and interaction, building and maintaining relationships, and encouraging teamwork (16).

The third component is the management of healthcare services, and our study again found a positive relationship between training and the development of clinical leadership. In their study, Dietz and Zwick (17) emphasised the role of workforce performance and service management in training and found that appropriately training staff can enhance organisational efficiency and produce additional benefits through increased employee productivity. The reason for this positive relationship is that effective leadership requires individuals capable of planning and managing elements such as data, performance and human resources while prioritising organisational success as the primary objective (16).

Regarding the fourth component, service improvement, the results of our study indicate a positive relationship between training and the development of this dimension of clinical leadership. Clarke et al. (18) have argued that in service quality improvement training, the curriculum serves as a valuable professional tool for organisations and has a significant impact on enhancing the quality of services provided. In a study by Taghadosi and Hossieni (19), it was noted that the lack of training—identified as a barrier to improving the quality of nursing services—could be addressed by implementing appropriate curriculum adjustments. This finding is consistent with the critical role of effective clinical leadership in enhancing patient safety and service quality (16).

Overall, the results of our study indicate a positive and significant relationship between the development of clinical leadership competencies and education. Kyriakidou et al. (20) have noted that participation in leadership training and development programmes positively influences healthcare professionals' leadership potential and clinical leadership characteristics. Gupta et al. (21) emphasised the widespread need for leadership and management training to develop clinical leadership skills. The study acknowledges the necessity of developing training content that aligns with the changing environment and society's needs for doctors. Organisations should support their doctors in developing their leadership skills, as new challenges may arise in the field of public health and healthcare. This support should include the allocation of sufficient time and resources to access education and targeted programmes (21). In their study on clinical leadership within the nursing curriculum, Brown et al. (22) noted that awareness of clinical leadership characteristics in healthcare could be raised among nursing students and other relevant stakeholders, and that training could be provided on how to utilise these characteristics. It was emphasised that healthcare services could be improved by selecting appropriate curriculum content and implementing effective educational approaches and strategies (22).

The results of our 13-hour clinical leadership training programme, in which participants demonstrated higher rates of correct responses, support the programme's contribution to learning outcomes. The scales used encompass the medical leadership competencies required to meet critical leadership needs in the Turkish healthcare environment. Our programme provided a targeted, multi-method, structured approach to developing management and leadership skills. A study by Lucas et al. (23) reported that leadership development programmes at medical schools in the US and Canada cover broad areas, such as leadership, collaboration and development, communication, guidance and managing change, business management, group dynamics and self-management. Similarly, the content of our programme aligns with the framework recommended for such programmes.

In our study, the correct response rate for participants who received training was approximately 74% in the pre-test and 61% in the post-test. However, correct response rates for leadership competencies

were higher among participants who received training than among those who did not. This situation may be related to the fact that the pre-test and post-test were not equivalent in content and level of difficulty.

We suggest that a 13-hour leadership training programme could serve as a appropriate starting point for imparting leadership knowledge and characteristics to healthcare professionals across all major specialisations. The increase in medical leadership competencies supports the future implementation of this programme for leadership training of doctors in Türkiye. Such programmes include pre- and post-assessment feedback and facilitate high-quality individual leadership performance within the healthcare sector. Frich et al.'s (6) review of 45 leadership training programmes from 1991 to 2013, reported that these programmes used various teaching methods and primarily focused on skill development and technical knowledge, while giving comparatively less attention to individual awareness and personal development. Furthermore, it was noted that the programmes primarily targeted junior doctors and middle managers and that the positive outcomes achieved were largely based on self-assessments (6). Pradarelli et al. (24) reported that a leadership training programme developed for surgical specialties positively impacted participants' leadership skills, self-awareness, and teamwork competencies. The use of both theoretical and experiential learning methods within the programme demonstrated that it facilitated the development of participants' knowledge and skills regarding leadership practices (24).

In a study based in Germany, Saravo et al. (25) conducted a four-week transformational and transactional leadership programme (the IMPACT leadership training module) for 57 junior doctors in a hospital setting and demonstrated that this short-term leadership training enhanced leadership skills among junior doctors.

In another study, Donaghy et al. (26) demonstrated improvements in all components examined during the expert trainers' leadership programme. It was found that this programme was beneficial not only to individual trainers but also to organisations (26). Within this scope, Kelly et al. (27) organised a five-week clinical leadership programme covering various leadership and management areas at a university hospital in Ireland. Steinert et al.'s (8) BEME systematic review (2012) confirms that faculty development initiatives focusing on leadership development in medical education have a positive impact on participants' attitudes, knowledge and skills. All participants reported that the programme was a positive learning experience. Furthermore, improvements were reported in leadership traits, business management and planning, career progression, and quality improvement. Straus et al. (28) reviewed the impact of clinical leadership programmes on doctors in academic health centres. They found that these programmes had modest effects on key outcomes and recommended that they be rigorously evaluated (28). Furthermore, they highlighted the potential of multidisciplinary clinical leadership programmes to strengthen understanding between working groups and enhance organisational leadership capabilities (29).

There is no standard approach to integrating clinical leadership training into pre-graduation education. However, in their study, Budakoglu et al. (30) reported that a course on leadership and teamwork integrated into the medical school curriculum increased students' leadership awareness and highlighted the need for development in this area.

Medical school curricula in Türkiye are believed not to allocate sufficient space for clinical leadership training. This situation may cause difficulties for doctors transitioning into leadership roles in clinical practice. The findings of this study support the necessity of developing clinical leadership training programmes tailored to healthcare professionals at different career stages. Furthermore, delivering such training via digital platforms to enhance accessibility could help reach a wider audience. This long-term, skill-focused clinical leadership curriculum could serve as a model or template for future studies and collaborations.

Study Limitations

This study has certain limitations. First, the study was conducted at a single centre and included only internal medicine residents, potentially limiting the generalisability of the findings. The snowball sampling method may have led to selection bias. Participants were not randomly assigned to groups (semi-experimental design), which limits the ability to draw causal inferences. A critical methodological limitation is that the parallel-form validity of the pre-test and post-test was not verified and the number of questions differed (pre-test: 14, post-test: 13). Consequently, intragroup pre- and post-test comparisons do not allow reliable interpretation. The primary outcome of the study should be evaluated by intergroup comparison in the post-test. Furthermore, leadership skills were assessed by knowledge levels and scale responses rather than by direct observation of clinical performance. The long-term effects of the training programme were not assessed. Therefore, evaluating knowledge retention through follow-up tests at 3 or 6 months is crucial for determining the sustainability of such educational interventions.

CONCLUSION

This study demonstrated that clinical leadership training improves leadership knowledge and awareness among internal medicine trainees. Participants who received training showed higher rates of correct responses than those who did not, supporting the programme's contribution to learning outcomes. However, the findings suggest the need for more comprehensive and continuous training programmes to develop leadership skills across all dimensions. Integrating clinical leadership training into the medical curriculum and continuing it at different career stages could contribute to the development of leadership capacity in healthcare.

Ethics

Ethics Committee Approval: Ethical approval for the study was obtained from the University of Health Sciences Türkiye, Ankara Bilkent City Hospital Ethics Committee no. 2 via decision no: E2-22-2526, dated: 28 September 2022.

Informed Consent: Informed consent was obtained from all participants prior to participation.

Acknowledgments

The authors would like to thank the internal medicine residents who participated in this study and the faculty members who supported the implementation of the clinical leadership training program.

Footnotes

Authorship Contributions

Concept: M.K., Ö.C., Design: M.K., Ö.C., Data Collection or Processing: M.K., Analysis or Interpretation: M.K., Ö.C., Writing: M.K., Ö.C.

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

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

REFERENCES

1. Daly J, Jackson D, Mannix J, Davidson PM, Hutchinson M. The importance of clinical leadership in the hospital setting. Journal of Healthcare Leadership. 2014; 75-83.
2. Swanwick T, McKimm J. What is clinical leadership…and why is it important? Clin Teach. 2011; 8: 22-6.
3. van Diggele C, Burgess A, Roberts C, Mellis C. Leadership in healthcare education. BMC Med Educ. 2020; 20: 456.
4. West M, Armit K, Loewenthal L, Eckert R, West T, Lee A. Leadership and leadership development in health care: the evidence base. London, Faculty of Medical Leadership and Management. © The Faculty of Medical Leadership and Management 2015.
5. Sadowski B, Cantrell S, Barelski A, O'Malley PG, Hartzell JD. Leadership training in graduate medical education: a systematic review. J Grad Med Educ. 2018; 10: 134-48.
6. Frich JC, Brewster AL, Cherlin EJ, Bradley EH. Leadership development programs for physicians: a systematic review. J Gen Intern Med. 2015; 30: 656-74.
7. Blumenthal DM, Bernard K, Bohnen J, Bohmer R. Addressing the leadership gap in medicine: residents' need for systematic leadership development training. Acad Med. 2012; 87: 513-22.
8. Steinert Y, Naismith L, Mann K. Faculty development initiatives designed to promote leadership in medical education. A BEME systematic review: BEME guide no. 19. Med Teach. 2012; 34: 483-503.
9. Anderson T, Shattuck J. Design-based research: a decade of progress in education research? Educational Researcher. 2012; 41: 16-25.
10. Budak F, Karasu F. Examination of the relationship between nursing students' intercultural sensitivity perceptions and clinical leadership characteristics. Journal of Healthcare Management and Leadership. 2020; 15-27.
11. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007; 39: 175-91.
12. NHS Institute for Innovation and Improvement. Clinical leadership competency framework. Coventry: NHS Institute for Innovation and Improvement; 2011.
13. Santa M, Nurcan S. Learning organization modelling patterns. Knowledge Management Research & Practice. 2016; 14: 106-25.
14. Hunt TM, Anderson M, Vo M, Penoyer DA. The impact of training on teamwork and simulated debriefings on real-life cardiopulmonary arrest events. Simulation & Gaming. 2021; 52: 585-600.
15. Harvey EM, Freeman D, Wright A, Bath J, Peters VK, Meadows G, et al. Impact of advanced nurse teamwork training on trauma team performance. Clinical Simulation in Nursing. 2019; 30: 7-15.
16. Academy of Medical Royal Colleges, NHS Institute for Innovation and Improvement. Medical leadership competency framework: enhancing engagement in medical leadership. 3rd ed. Coventry: NHS Institute for Innovation and Improvement; 2010.
17. Dietz D, Zwick T. The retention effect of training: portability, visibility, and credibility. The International Journal of Human Resource Management. 2021; 33: 710-41.
18. Clarke MJ, Steffens FL, Mallory GW, Starr SR, Porter BL, Krauss WE, et al. Incorporating quality improvement into resident education: structured curriculum, evaluation, and quality improvement projects. World Neurosurg. 2019; 126: e1112-20.
19. Taghadosi M, Hossieni F. Challenges of nursing quality improvement in Iran: a systematic review. Scientific Journal of Nursing, Midwifery and Paramedical Faculty. 2018; 4: 6-16.
20. Kyriakidou N, Aspasia G, George P, Anastasios S, Marios A. Leadership development in health care: the role of clinical leaders. Journal of Human Resource and Sustainability Studies. 2021; 9: 231-49.
21. Gupta AK, Neely J, Wilkinson C. Clinical leadership training: a clinician's perspective. Clin Radiol. 2021; 76: 241-6.
22. Brown A, Dewing J, Crookes P. Clinical leadership as an integral curriculum thread in pre-registration nursing programmes. Nurse Educ Today. 2016; 38: 9-14.
23. Lucas R, Goldman EF, Scott AR, Dandar V. Leadership development programs at academic health centers: results of a national survey. Acad Med. 2018; 93: 229-36.
24. Pradarelli JC, Jaffe GA, Lemak CH, Mulholland MW, Dimick JB. A leadership development program for surgeons: first-year participant evaluation. Surgery. 2016; 160: 255-63.
25. Saravo B, Netzel J, Kiesewetter J. The need for strong clinical leaders - transformational and transactional leadership as a framework for resident leadership training. PLoS One. 2017; 12: e0183019.
26. Donaghy G, McKeever K, Flanagan C, O'Kane D, McQuillan B, Cash J, et al. Helping doctors in training to STEP-UP: a leadership and quality improvement programme in the belfast health and social care trust. Ulster Med J. 2018; 87: 112-6.
27. Kelly D, McErlean S, Naff K. Outcomes of a clinical leadership training program amongst hospital doctors. Ir Med J. 2018; 111: 733.
28. Straus SE, Soobiah C, Levinson W. The impact of leadership training programs on physicians in academic medical centers: a systematic review. Acad Med. 2013; 88: 710-23.
29. Frich JC, Spehar I. Physician leadership development: towards multidisciplinary programs? BMJ Leader. 2018; 2: 91-4.
30. Budakoglu II, Coskun O, Karabacak O, Karabacak NI. Curriculum implementation for leadership and teamwork: medical students. GMJ. 2021; 32: 23-7.

Gazi Med J 2026;37(4):518-523

DOI: http://dx.doi.org/10.12996/gmj.2026.4747

Spectrum of *CYP21A2* Variants: A Retrospective Single-center Study from a Tertiary Healthcare Center in Türkiye

Üçüncü Basamak Sağlık Merkezi Genetik Laboratuvarında *CYP21A2* Genine Yönelik Test Yapılan Olguların Retrospektif Olarak Değerlendirilmesi

Ahmet Kablan[1], Furkan Baştan[1], Aysu Yeşim Tezcan[2], Asuman Bölükbaşı[3], Şenay Savaş Erdeve[4], Haktan Bağış Erdem[1]

[1]Department of Medical Genetics, University of Health Sciences Türkiye, Ankara Etlik City Hospital, Ankara, Türkiye
[2]Department of Women's Health and Obstetrics, University of Health Sciences Türkiye, Ankara Etlik City Hospital, Ankara, Türkiye
[3]Techno Health Laboratory & Diagnostic Health Services Industry and Trade, İstanbul, Türkiye
[4]Department of Paediatric Endocrinology, University of Health Sciences Türkiye, Ankara Etlik City Hospital, Ankara, Türkiye

ABSTRACT

Objective: Congenital adrenal hyperplasia due to 21-hydroxylase deficiency is a common autosomal recessive disorder caused by pathogenic variants in the *CYP21A2* gene. The spectrum of variants differs significantly among populations because of complex genetic recombination events involving the neighbouring *CYP21A1P* pseudogene. This study aimed to evaluate the variant distribution, genotype frequencies, and clinical indications in a large Turkish cohort referred for *CYP21A2* genetic analysis.

Methods: A total of 252 pediatric patients who underwent *CYP21A2* gene sequencing and multiplex ligation-dependent probe amplification at a single tertiary center were retrospectively analyzed. Genotype and allele frequencies of classical point variants, small indels, large deletions, and pseudogene-derived micro-conversions were calculated and correlated with clinical referral indications.

Results: Among the 252 patients, 133 (52.8%) were, by both sequencing and copy number analyses, found to carry no reportable variants. The most frequent pathogenic variant identified was c.844G>T p.(Val282Leu), found in the homozygous state in 20 patients (7.9%) and in 77 alleles in total. Other prevalent, severe variants included the intron 2 splice- site variant c.293-13A/C>G and the pseudogene-derived p.(Gln319*) variant. Structural rearrangements, including exon-specific micro-conversions (most notably involving exon 4), large multi-exon deletions, and complex allele combinations such

ÖZ

Amaç: 21-hidroksilaz eksikliğine bağlı konjenital adrenal hiperplazi, *CYP21A2* genindeki patojenik varyantların neden olduğu, sık görülen bir otozomal resesif hastalıktır. Mutasyon spektrumu, komşu *CYP21A1P* psödogeni ile meydana gelen karmaşık genetik rekombinasyon olayları nedeniyle popülasyonlar arasında önemli ölçüde farklılık göstermektedir. Bu çalışma, *CYP21A2* genetik analizi için yönlendirilen geniş bir Türk kohortundaki varyant dağılımını, genotip frekanslarını ve klinik endikasyonları değerlendirmeyi amaçlamıştır.

Yöntemler: Tek bir üçüncü basamak merkezde *CYP21A2* gen dizilemesi ve çoklu ligasyon bağımlı prob amplifikasyonu uygulanan toplam 252 pediyatrik hasta geriye dönük (retrospektif) olarak analiz edilmiştir. Klasik nokta mutasyonlarının, küçük indellerin (insersiyon/delesyon), büyük delesyonların ve psödogen kaynaklı mikro-konversiyonların genotip ve alel frekansları hesaplanmış ve klinik sevk endikasyonları ile ilişkilendirilmiştir.

Bulgular: Toplam 252 hasta arasında 133'ü (%52,8) hem dizileme hem de kopya sayısı analizleri için raporlanabilir bir varyanta sahip değil olarak saptanmıştır. Tespit edilen en sık patojenik varyant, 20 hastada (%7,9) homozigot durumda ve toplamda 77 alelde bulunan c.844G>T p.(Val282Leu) varyantı olmuştur. Diğer yaygın varyantlar arasında intron 2 splicing bölgesi mutasyonu c.293-13A/C>G ve psödogen kaynaklı p.(Gln319*) varyantı yer almıştır. En belirgin şekilde ekzon 4'ü içeren ekzona özgü mikro-konversiyonlar ve büyük çoklu ekzon

Cite this article as: Kablan A, Baştan F, Tezcan AY, Bölükbaşı A, Savaş Erdeve Ş, Erdem HB. Spectrum of *CYP21A2* variants: a retrospective single-center study from a tertiary healthcare center in Türkiye. Gazi Med J. 2026;37(4):518-523

Address for Correspondence/Yazışma Adresi: Ahmet Kablan, Department of Medical Genetics, University of Health Sciences Türkiye, Ankara Etlik City Hospital, Ankara, Türkiye
E-mail / E-posta: kablanmd@gmail.com
ORCID ID: orcid.org/0000-0002-8966-2208

Received/Geliş Tarihi: 23.06.2026
Accepted/Kabul Tarihi: 25.08.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

as CAH-X, were detected in 21 patients (8.3%). Hirsutism, abnormal menstruation, and precocious puberty were the leading indications for genetic testing in adolescent cohorts, whereas scrotal/genital hyperpigmentation and abnormal neonatal screening predominated in infancy.

Conclusion: The genetic architecture of *CYP21A2* in Türkiye is characterized by a high prevalence of the non-classic c.844G>T p.(Val282Leu) variant alongside a diverse spectrum of severe pseudogene-mediated conversions and deletions. These findings underscore the clinical necessity of using parallel sequencing and copy number detection strategies to achieve accurate molecular diagnosis and genetic counselling in the Turkish population.

Keywords: Congenital adrenal hyperplasia, *CYP21A2* variants, population genetics, Turkish population

ÖZ

delesyonları dahil olmak üzere yapısal düzenlemeler, CAH-X gibi karmaşık alel kombinasyonları da dahil olmak üzere 21 hastada (%8,3) tespit edilmiştir. Hirsutizm, anormal menstrüasyon (adet düzensizliği) ve prekoks puberte (erken ergenlik), kohortta ileri yaş grubu için genetik test için önde gelen endikasyonlar iken; skrotal/genital hiperpigmentasyon ve anormal yenidoğan taraması süt çocukluğu (infant) döneminde baskın olmuştur.

Sonuç: Türkiye'deki *CYP21A2* genetik yapısı, klasik olmayan c.844G>T p.(Val282Leu) varyantının yüksek prevalansının yanı sıra, şiddetli psödogen aracılı konversiyonlar ve delesyonlardan oluşan çeşitli bir spektrum ile karakterizedir. Bu bulgular, Türk popülasyonunda doğru moleküler tanı ve genetik danışmanlık sağlamak için paralel dizileme ve kopya sayısı tespit stratejilerinin birlikte kullanılmasının klinik gerekliliğini vurgulamaktadır.

Anahtar Sözcükler: Konjenital adrenal hiperplazi, *CYP21A2* varyantları, popülasyon genetiği, Türk popülasyonu

INTRODUCTION

Congenital adrenal hyperplasia (CAH) encompasses a group of autosomal recessive enzymatic disorders characterized by impaired steroid hormone biosynthesis in the adrenal cortex. Over 95% of CAH cases result from 21-hydroxylase deficiency, which leads to a systemic accumulation of cortisol precursors (such as 17-hydroxyprogesterone) and a subsequent shift toward excessive adrenal androgen production (1). Depending on the residual enzymatic activity, the clinical spectrum ranges from severe, life-threatening classic forms (salt-wasting and simple virilizing) manifesting in infancy to milder, non-classic (late-onset) forms that present during childhood or adolescence with signs of hyperandrogenism, premature adrenarche, hirsutism, or menstrual irregularities (2).

The *CYP21A2* gene, which encodes the active 21-hydroxylase enzyme, is located within the highly complex major histocompatibility complex class III region on chromosome 6p21.3 (3). It spans approximately 3.4 kb across 10 exons and lies adjacent to a highly homologous, non-functional pseudogene, *CYP21A1P*, with which it shares approximately 98% exonic sequence identity, complicating testing and analysis. This tandem arrangement and structural homology predispose the locus to frequent intergenic recombination events. These include unequal crossing-over leading to large deletions or duplications (such as the CAH-X chimeric gene fusion with *TNXB*) and gene conversion events that transfer deleterious point variants from the inactive pseudogene to the active *CYP21A2* locus (4).

More than 200 pathogenic variants have been described in *CYP21A2*, although most disease-causing variants originate from a limited number of deleterious sequences transferred from the highly homologous pseudogene *CYP21A1P* through gene conversion and unequal recombination events. Large 30-kb deletions account for approximately 20-30% of classical CAH alleles and are typically associated with severe loss of 21-hydroxylase activity, although breakpoint location may occasionally result in milder phenotypes (5). Most affected individuals are compound heterozygotes carrying different variants on each allele, with disease severity generally determined by the less severe variant. In addition, approximately 10% of patients harbor CAH-X syndrome, a contiguous gene deletion involving both *CYP21A2* and *TNXB*, which combines features of CAH with connective tissue manifestations such as joint hypermobility, recurrent dislocations, hernias, and cardiovascular abnormalities (6).

Data regarding the comprehensive spectrum of *CYP21A2* alterations in Turkish populations, particularly in pediatric cohorts, remain limited. Given the marked genetic heterogeneity of the *CYP21A2* locus and the clinical importance of accurate molecular diagnosis, population-specific data are essential for improving diagnostic strategies. Therefore, this study aimed to characterize the distribution of disease-causing *CYP21A2* variants and structural rearrangements in a large referral cohort of 252 pediatric subjects evaluated at a tertiary medical genetics center in Türkiye, and to assess the additional diagnostic value of multiplex ligation-dependent probe amplification (MLPA) used alongside conventional Sanger sequencing. Characterization of the variational landscape in such large clinical series may contribute to the optimization of molecular diagnostic algorithms and future screening approaches.

MATERIALS AND METHODS

Study Population

This retrospective cohort study evaluated 252 pediatric patients who underwent molecular genetic screening for suspected 21-hydroxylase deficiency before the age of 18 (when the testing procedure was implemented) at the Medical Genetics Laboratory of University of Health Sciences Türkiye, Ankara Etlik City Hospital between 2022 and 2026. Patients were referred for a wide range of indications, including premature adrenarche or pubarche, precocious puberty, hirsutism, severe acne, polycystic ovary syndrome (PCOS) refractory to standard management, ambiguous genitalia, scrotal or labial hyperpigmentation, unexplained neonatal adrenal insufficiency, and positive or borderline neonatal screening results. Patients were divided into seven categories based on these symptoms: Definitive clinical diagnosis; disorders of sexual development' (ambiguous genitalia, scrotal/labial hyperpigmentation); puberty-related problems (menstrual problems, premature adrenarche or

pubarche, precocious puberty, hirsutism, severe acne, PCOS, short stature); fertility problems; abnormal screening tests; adrenal insufficiency; and family history. Ethics committee approval for the study was received from the Scientific Research Evaluation and Ethics Committee of University of Health Sciences Türkiye, Ankara Etlik City Hospital (approval number: AEŞH-BADEK1 2026-434/date: 06.05.2026). Informed consent forms for publication were obtained from the adult patients and from the legal guardians of the pediatric patients.

Genetic Analysis

Genomic DNA was isolated from peripheral blood samples using standard protocols. Following amplification, the 10 coding exons and flanking intronic regions were sequenced using automated Sanger sequencing (Seqline-EYS Medikal, İstanbul, Türkiye/ABI 3500xl DNA Analyzer). Sequence annotations were completed relative to the reference sequence NM_000500.5. To complement sequencing and systematically evaluate copy-number variations, large deletions, duplications, and gene conversions, MLPA was performed using the SALSA MLPA P050 kit (MRC Holland, Amsterdam, Netherlands). Detected variations were curated and classified according to the joint consensus guidelines of the American College of Medical Genetics and Genomics (7). Patients with biallelic compound variants and complex rearrangements, for whom parental Sanger sequencing and MLPA analyses could not confirm that the variants were in trans or were causative, were excluded from the study. Conversions were reported only if the findings were consistent across Sanger sequencing and MLPA analysis and if the results of the two tests correlated.

Statistical Analysis

Descriptive statistics were applied to summarize demographic features, genotype frequencies, and allele occurrences. Quantitative metrics are expressed as means and interquartile ranges, whereas categorical parameters are presented as absolute counts and percentages. This study focused mainly on genetic frequency profiles, excluding biochemical and clinical parameters. Cumulative variant prevalence was evaluated using reference genomic data sourced from three primary repositories: The Turkish Genome Project (TGP, v3.1; accession: June 2026), the Turkish Variome on the Franklin Genoox platform (accession: June 2026), and the Genome Aggregation Database (gnomAD v4.1.1; accession: June 2026).

RESULTS

Demographic and Clinical Characteristics

The study comprised 252 individuals aged 0 months (neonates) to 18 years. The sex distribution showed a female predominance: 150 females (59.5%) versus 102 males (40.5%), reflecting the broader clinical presentation of non-classic phenotypes during adolescence and young adulthood. The median age was 7.9 years (interquartile range, 0-13 years). Clinical referral indications varied markedly by age group: adolescent female patients presented primarily with hirsutism, abnormal menstruation, acne, or PCOS, whereas pediatric patients presented with precocious puberty, premature pubarche/adrenarche, or short stature. Infants and neonates presented with scrotal or labial hyperpigmentation, ambiguous genitalia, clinical adrenal crisis, or abnormal neonatal dried-blood spot screening tests.

Regarding referral indications, an abnormal screening test was the primary reason for clinical admission, accounting for 23.8% (n = 60) of the cohort and substantially contributing to the high proportion of infant evaluations. Among adolescent and adult presentations, hyperandrogenism-related symptoms predominated; hirsutism was documented in 17% (n = 43) of all patients, and menstrual irregularities (abnormal menstruation) were reported in 12.7% (n = 32). A preliminary clinical diagnosis of CAH, requested by the referring physicians and based on consistent clinical and biochemical findings, was the indication in 13.2% (n = 34) of cases. Other less frequent referral pathways included a positive family history, PCOS phenotypes, isolated short stature, and overt adrenal insufficiency (Supplementary Material).

Genetic Findings

The analysis revealed a broad spectrum of molecular variants; however, no diagnostic pathogenic or likely pathogenic alterations were detected in a significant proportion of patients (n = 133, 52.8%). Among the molecularly confirmed cases, the most prevalent point variant was the c.844G>T p.(Val282Leu), detected in 77 alleles and accounting for a cohort allele frequency of 0.152. This variant was observed in both homozygous and heterozygous states (isolated or compound) across various clinical indications. Other frequently observed pathogenic variants included the intron 2 splice-site variant c.293-13A/C>G (n = 25, allele frequency: 0.049) and the missense variant c.1360C>T p.(Pro454Ser) (n = 13, allele frequency: 0.025). Notably, severe classical variants were identified in patients exhibiting classical CAH phenotypes: The nonsense variant c.955C>T p.(Gln319*) was found in 11 alleles (allele frequency: 0.021), while the missense alterations c.392G>A p.(Gly131Asp) (n = 8, allele frequency: 0.015) and c.518T>A p.(Ile173Asn) (n = 5, allele frequency: 0.009) were well represented in the classical presentation subgroup. In 43 patients, only one disease-causing variant was detected. In addition to pathogenic or likely pathogenic variants, 4 variants of unknown significance that were thought to be causative (as a result of phasing, biochemical profile, etc.) were reported.

Beyond single nucleotide variants, MLPA analysis identified complex structural rearrangements and copy number variations in 21 individuals. These included heterozygous or homozygous micro-conversions of exon 4, exon 4-8 and exon 4-9 micro-conversions, and various multi-exon deletions (e.g., deletion of exons 1-3). Heterozygous or homozygous micro-conversions in exon 4 were identified in 12 alleles, corresponding to a cohort frequency of 0.023. Complete gene deletions, large homozygous deletions characterizing the CAH-X syndrome, and total gene conversion events were also identified in a small subset of subjects (Supplementary Material).

DISCUSSION

This study provides a comprehensive overview of *CYP21A2* sequence variants and structural rearrangements in a Turkish pediatric referral cohort. The findings demonstrate genetic heterogeneity and confirm the importance of using complementary molecular techniques for accurate diagnosis.

The predominance of p.(Val282Leu) observed in this cohort is consistent with its recognized status as one of the most common *CYP21A2* variants associated with milder or non-classical forms of CAH. The frequent detection of this variant among individuals presenting with hirsutism, menstrual irregularities, and hyperandrogenic symptoms highlights its clinical relevance in endocrine referral populations. This variant has traditionally been associated with the non-classical form of CAH and is known to retain partial 21-hydroxylase activity, providing a biochemical basis for delayed diagnosis. This finding was consistently observed in our cohort, where carriers of this variant predominantly presented with milder, later-onset clinical manifestations. Consequently, as shown in Table 1, this variant exhibits a considerably higher frequency in population genomic databases compared with most other disease-causing *CYP21A2* alleles. Consistent with its relatively high carrier frequency, homozygosity for this variant was also observed in a substantial proportion of affected individuals within our cohort.

The recurrent identification of c.293-13A/C>G, p.(Gln319*), p.(Pro454Ser) further highlights the strong predominance of well-established pathogenic *CYP21A2* variants within our cohort. Overall, our findings demonstrate a clear dominance of a limited set of approximately ten classical pathogenic variants that have been repeatedly reported in previous studies and are known to account for a substantial proportion of disease alleles in CAH populations (8). These variants have been consistently associated with markedly reduced enzymatic activity and are typically enriched among patients with classical forms of CAH.

In addition to this well-defined variant spectrum, we identified several rare variants that are currently classified as variants of uncertain significance according to ACMG criteria and may represent population-specific alleles or reflect enrichment in the Turkish cohort, such as c.865C>T p.(Leu289Phe), c.1352C>T p.(Thr451Met), c.86A>C p.(His29Pro). To date, there is limited large-scale clinical or functional evidence regarding the pathogenicity and genotype-phenotype correlations of these variants, and they are either absent from or only sparsely represented in genome databases (Table 1). Therefore, further investigation in larger, well-characterized cohorts, ideally supported by functional studies, will be essential to clarify their clinical relevance and refine genotype-phenotype interpretation in CAH.

Compared with neighbouring countries, the variant spectrum appears to resemble that of western neighbours more closely

Table 1. Allele frequencies of identified CYP21A2 variants compared with reference datasets.

Variant (NM_000500.5)	Allele count (total n=504)	Cohort allele frequency	TGP frequency	TV frequency	gnomAD frequency
c.844G>T p.(Val282Leu)	77	0.152	0.018/14*	-	0.005515/2*
c.293-13A/C>G	25	0.049	**0.0029**	-	0.002420/7*
c.1360C>T p.(Pro454Ser)	13	0.025	0.0059/2*	-	**0.006282/3***
c.955C>T p.(Gln319*)	11	0.021	**0.012**	-	**0.0008961**
c.392G>A p.(Gly131Asp)	8	0.015	**0.000097**	-	0
c.518T>A p.(Ile173Asn)	5	0.009	**0.00058**	-	0.0005579/3*
c.1447C>T p.(Pro483Ser)	5	0.009	**0.0026**	**0.0022**	0.0004815/20*
c.92C>T p.(Pro31Leu)	4	0.0079	**0.00049**	-	0.0001412/2*
c.961G>A p.(Glu321Lys)	2	0.0039	-	-	-
c.1069C>T p.(Arg357Trp)	5	0.009	**0.000097**	-	**0.00003537**
c.1019G>A p.(Arg340His)	2	0.0039	**0.00068**	0.001537	**0.000009918**
c.850A>G p.(Met284Val)	2	0.0039	**0.000097**	-	**0.00004836**
c.865C>T p.(Leu289Phe)	2	0.0039	**0.000097**	-	**0.000009921**
c.1352C>T p.(Thr451Met)	1	0.0019	-	-	**0.00004909**
c.397C>T p.(Arg133Cys)	1	0.0019	**0.000097**	0.000377	**0.00006701/1***
c.86A>C p.(His29Pro)	1	0.0019	-	-	0.00000062
c.1174G>A p.(Ala392Thr)	1	0.0019	**0.00078**	-	**0.007159**
c.188A>T p.(His63Leu)	1	0.0019	**0.000097**	-	**0.02901**

*: Number of homozygous individuals. Structural changes were excluded, TGP: Turkish Genome Project, TV: Turkish Variome.

than that of eastern neighbours. Total sample sizes vary across studies, which may influence these observations. Moreover, when reports from geographically similar populations in the literature are collectively evaluated, a consistent theme is the presence of a distinct, country-specific genetic landscape, underscoring the unique population-specific structure of each cohort (9,10) .

Overall, the detection rate of biallelic pathogenic variants was significantly higher in both the clinically compatible diagnosis group and the puberty-related disorders group, with a cumulative detection rate of 36.8% (35/95). In contrast, the molecular diagnostic yield was substantially lower in groups presenting with isolated menstrual problems or hirsutism and in those referred because of suspicious screening test results, with detection rates of 16.4% (11/67) and 3.3% (2/60), respectively. Among molecularly confirmed cases, monoallelic heterozygous variants were detected in 17.1% of patients, whereas biallelic alterations—comprising homozygous (16.7%) and compound heterozygous (13.5%) genotypes—comprised the remainder of the cohort (Table 2).

CAH is included in the neonatal screening program in Türkiye, and, correspondingly, an abnormal screening result constitutes a substantial proportion of the indication criteria for inclusion in our cohort (23.8%, n = 60). Additionally, the higher frequency of females in the cohort may be attributed to shunting of accumulated biochemical materials toward androgen production in cases of 21-hydroxylase deficiency. This leads to more frequent and more pronounced clinical presentations in females and naturally drives these individuals to seek medical attention.

Table 2. Cumulative distribution of *CYP21A2* genotypes according to their zygosity in the cohort.

Genotype (Sanger sequence/MLPA)	Number of patients (n)	Percentage (%)
Wild-type (no pathogenic variant detected)	133	52.8
Monoallelic genotype [heterozygous only]	43	17.1
Biallelic genotype [homozygous]	42	16.7
Biallelic genotype [compound heterozygous]	34	13.5
Total	252	100

MLPA: Multiplex ligation-dependent probe amplification.

Study Limitations

This study has several limitations. First, because this study was a retrospective evaluation conducted at a single tertiary referral center, the cohort may be subject to referral bias and may not fully reflect the general Turkish population. The highly complex genetic architecture of this region, particularly the presence of pathogenic variants involving structural rearrangements, may not always be adequately detected by Sanger sequencing combined with MLPA analysis. In this context, long-read sequencing approaches are likely to provide additional clinical benefit. Additionally, variants considered to have no clear clinical impact—such as gene duplications, exon duplications, or combinations like p.Gln319* with heterozygous total duplication—which are frequently encountered in analysis were not included in the frequency analyses, as the study was primarily focused on clinically oriented pathogenic variants (5). Because national genomic databases included a limited number of alleles at the time of this analysis, larger multicenter studies are required to establish definitive nationwide carrier frequencies for rare and non-classical structural variants.

CONCLUSION

The *CYP21A2* variant landscape in Türkiye is characterized by a high prevalence of the non-classic c.844G>T p.(Val282Leu) variant and a diverse spectrum of severe pseudogene-mediated micro-conversions and macro-deletions. These findings emphasize the clinical importance of combining sequence analysis with copy number methodologies to ensure accurate molecular diagnosis, optimize clinical management, and guide genetic counseling for affected families.

Ethics

Ethics Committee Approval: Ethics committee approval for the study was received from the Scientific Research Evaluation and Ethics Committee of University of Health Sciences Türkiye, Ankara Etlik City Hospital (approval number: AEŞH-BADEK1 2026-434/date: 06.05.2026).

Informed Consent: Informed consent forms for publication were obtained from the adult patients and from the legal guardians of the pediatric patients.

Footnotes

Authorship Contributions

Concept: A.K., Design: A.K., Data Collection or Processing: A.K., F.B., A.B., H.B.E., Ş.S.E., A.Y.T., Analysis or Interpretation: A.K., F.B., H.B.E., Ş.S.E., A.B., Literature Search: A.K., Writing: A.K., F.B.

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

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

Supplementary Material Link: https://d2v96fxpocvxx.cloudfront.net/8a9ff4da-541a-42fa-9980-1a9a3ab6d6c5/content-images/359b49c1-c9e8-4be8-9fc5-97956e1516ce.pdf

REFERENCES

1. Krone N, Dhir V, Ivison HE, Arlt W. Congenital adrenal hyperplasia and P450 oxidoreductase deficiency. Clin Endocrinol (Oxf). 2007; 66: 162-72.
2. Speiser PW, White PC. Congenital adrenal hyperplasia. N Engl J Med. 2003; 349: 776-88.
3. Carroll MC, Campbell RD, Porter RR. Mapping of steroid 21-hydroxylase genes adjacent to complement component C4 genes in HLA, the major histocompatibility complex in man. Proc Natl Acad Sci U S A. 1985; 82: 521-5.
4. Merke DP, Chen W, Morissette R, Xu Z, Van Ryzin C, Sachdev V, et al. Tenascin-X haploinsufficiency associated with Ehlers-Danlos syndrome in patients with congenital adrenal hyperplasia. J Clin Endocrinol Metab. 2013; 98: E379-87.

5. Concolino P, Costella A. Congenital adrenal hyperplasia (CAH) due to 21-hydroxylase deficiency: a comprehensive focus on 233 pathogenic variants of CYP21A2 gene. Mol Diagn Ther. 2018; 22: 261-80.
6. Merke DP, Auchus RJ. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. N Engl J Med. 2020; 383: 1248-61.
7. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al.; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015; 17: 405-24.
8. Baumgartner-Parzer S, Witsch-Baumgartner M, Hoeppner W. EMQN best practice guidelines for molecular genetic testing and reporting of 21-hydroxylase deficiency. Eur J Hum Genet. 2020; 28: 1341-67.
9. Fylaktou I, Mertzanian A, Farakla I, Gryparis A, Vasilakis IA, Binou M, et al. Genetics of 21-OH deficiency and genotype-phenotype correlation: experience of the Hellenic National Referral Center. Curr Issues Mol Biol. 2024; 46: 10696-713.
10. Rabbani B, Mahdieh N, Ashtiani MT, Larijani B, Akbari MT, New M, et al. Mutation analysis of the *CYP21A2* gene in the Iranian population. Genet Test Mol Biomarkers. 2012; 16: 82-90.

Gazi Med J 2026;37(4):524-532

DOI: http://dx.doi.org/10.12996/gmj.2026.4750

Submandibular Gland Involvement in Head and Neck Cancers: A Single-center Retrospective Study of 175 Cases

Baş ve Boyun Kanserlerinde Submandibular Bez Tutulumu: 175 Olguluk Tek Merkezli Retrospektif Çalışma

Muhittin Akalın[1], Mehmet Ekrem Zorlu[2], Emirhan Akyol[3], Betül Öğüt[4], Utku Aydil[2], Metin Yılmaz[2], Mehmet Düzlü[2]

[1]Department of Otolaryngology and Head & Neck Surgery, University of Health Sciences Türkiye, Ankara Bilkent City Hospital, Ankara, Türkiye
[2]Department of Otolaryngology and Head & Neck Surgery, Gazi University, Faculty of Medicine, Ankara, Türkiye
[3]Clinic of Otorhinolaryngology and Head & Neck Surgery, Adıyaman Training and Research Hospital, Adıyaman, Türkiye
[4]Department of Pathology, Gazi University, Faculty of Medicine, Ankara, Türkiye

ABSTRACT

Objective: The submandibular gland is routinely removed during level Ib neck dissection for head and neck cancer. This study evaluated the feasibility of gland preservation in appropriate patients considering oncological risks.

Methods: This retrospective study evaluated 175 patients with head and neck cancer who underwent neck dissection including Level Ib between 2015 and 2024. Level Ib nodal status was assessed in all patients, whereas submandibular gland involvement was evaluated in patients with at least one excised gland (n = 165). Overall survival (OS) and recurrence-free survival (RFS) were compared between gland-sparing and non-gland-sparing surgery in the oral cavity squamous cell carcinoma (SCC) cohort (n = 125) using the Kaplan-Meier method and log-rank test.

Results: The predominant histopathological type was SCC (89.1%), and the oral cavity was the most common tumor site (72.0%). Submandibular gland involvement was observed in 1 of 165 patients (0.6%). In the oral cavity SCC cohort, no statistically significant differences were found in OS (p = 0.288) or RFS (p = 0.936) between gland-sparing surgery (n = 12) and non-gland-sparing surgery (n = 113).

Conclusion: Submandibular gland involvement is rare, and gland-sparing surgery may be viable in carefully selected oral cavity SCC patients. Larger studies are required to clarify oncological reliability

ÖZ

Amaç: Submandibular bez, baş ve boyun kanserlerinde Level Ib boyun diseksiyonu sırasında rutin olarak çıkarılmaktadır. Bu çalışmada, onkolojik riskler göz önünde bulundurularak uygun hastalarda submandibular bezin korunmasının uygulanabilirliğinin değerlendirilmesi amaçlandı.

Yöntemler: Bu retrospektif çalışmada, 2015-2024 yılları arasında Level Ib'yi içeren boyun diseksiyonu uygulanan baş ve boyun kanserli 175 hasta değerlendirildi. Tüm hastalarda Level Ib lenf nodlarının durumu değerlendirildi; submandibular bez tutulumu ise en az bir submandibular bezi eksize edilen hastalarda (n = 165) incelendi. Glandüler koruyucu ve bez koruyucu olmayan cerrahi uygulanan hastalar arasındaki genel sağkalım (OS) ve nükssüz sağkalım (RFS), oral kavite skuamöz hücreli karsinom (SCC) kohortunda (n = 125) Kaplan-Meier yöntemi ve log-rank testi kullanılarak karşılaştırıldı.

Bulgular: En sık görülen histopatolojik tip SCC (%89,1) olup, en yaygın tümör lokalizasyonu oral kaviteydi (%72,0). Submandibular bez tutulumu 165 hastanın 1'inde (%0,6) saptandı. Oral kavite SCC kohortunda, bez koruyucu cerrahi (n = 12) ile glandüler koruyucu olmayan cerrahi (n = 113) arasında OS (p = 0,288) veya RFS (p = 0,936) açısından istatistiksel olarak anlamlı fark saptanmadı.

Sonuç: Submandibular bez tutulumu nadirdir ve dikkatle seçilmiş oral kavite SCC hastalarında submandibular bezin korunmasına yönelik

Cite this article as: Akalın M, Zorlu ME, Akyol E, Öğüt B, Aydil U, Yılmaz M, et al. Submandibular gland involvement in head and neck cancers: a single-center retrospective study of 175 cases. Gazi Med J. 2026;37(4):524-532

Address for Correspondence/Yazışma Adresi: Muhittin Akalın, Department of Otolaryngology and Head & Neck Surgery, University of Health Sciences Türkiye, Ankara Bilkent City Hospital, Ankara, Türkiye
E-mail / E-posta: muhittin.akalin@saglik.gov.tr
ORCID ID: orcid.org/0000-0002-9402-6943

Received/Geliş Tarihi: 25.06.2026
Accepted/Kabul Tarihi: 21.09.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

and functional benefits and support patient selection criteria for preservation.

Keywords: Head and neck neoplasms, submandibular gland, neck dissection, squamous cell carcinoma of head and neck, mouth neoplasms, lymphatic metastasis

ÖZ

cerrahi uygulanabilir bir seçenek olabilir. Onkolojik güvenilirliği ve fonksiyonel yararları daha net ortaya koymak ve bez koruma için hasta seçim kriterlerini belirlemek amacıyla daha geniş kapsamlı çalışmalara ihtiyaç vardır.

Anahtar Sözcükler: Baş ve boyun neoplazmları, submandibular bez, boyun diseksiyonu, baş ve boyun skuamöz hücreli karsinomu, ağız neoplazmları, lenfatik metastaz

INTRODUCTION

Neck dissection is a major component of the treatment of head and neck cancers. Since the definition of radical neck dissection by Crile (1), many changes have occurred in neck dissections, such as modified radical neck dissection and selective neck dissections (1-3).

Most head and neck cancers carry at least an intermediate risk of cervical metastasis (4). Neck metastasis can be diagnosed through physical examination or imaging techniques. The prognosis of head and neck cancers is directly related to neck involvement (5-7). In general, a risk of occult metastasis greater than 20% is considered an indication for elective neck dissection (8). Selective or functional neck dissection, rather than comprehensive neck dissection (modified radical or radical neck dissection), is more appropriate for elective neck dissection (9-11).

Standard level Ib dissection, including the excision of the submandibular gland (SMG) with the surrounding lymph nodes and adipose tissue, is part of elective or comprehensive neck dissections (9,12). Submandibular glands do not contain intraparenchymal lymph nodes, unlike the parotid gland, which includes an average of 10 to 20 intraglandular lymph nodes (13). This raises the question: "Is it feasible to perform level Ib neck dissection while preserving the SMG in selected cases without compromising oncological outcomes?"

The SMG is one of the major salivary glands responsible for salivary function. Removal of a single SMG usually does not cause clinically significant xerostomia. However, bilateral dissection or adjuvant radiotherapy may increase the risk of this complication (14). In addition, there is a low risk of hypoglossal and lingual nerve paralysis complications after SMG excision (15).

Therefore, in this study, we aimed to evaluate SMG involvement in routine histopathological analyses of neck dissection specimens obtained from patients with head and neck cancers. In addition, oncological outcomes were compared between gland-sparing and non-gland-sparing surgery in the oral cavity squamous cell carcinoma (SCC) cohort, in which gland-sparing procedures were performed.

MATERIALS AND METHODS

This study was a retrospective analysis of patients who underwent neck dissection, including level Ib, for head and neck cancer at a single tertiary referral center between January 1, 2015, and January 1, 2024. All patients were evaluated by a surgical team consisting of otolaryngology specialists, and surgery was performed by a senior head and neck surgeon.

The exclusion criteria were as follows:

1. Patients who had previously been treated at our center or another institution and had developed SCC recurrence,

2. Patients with incomplete clinical data,

3. Cases in which primary treatment was administered using a non-surgical method.

A total of 175 patients who underwent neck dissection including level Ib were included, and level Ib nodal status was assessed in all cases. SMG involvement could be evaluated in patients in whom at least one SMG was excised (n = 165); the remaining 10 patients had no SMG submitted for histopathologic assessment because of bilateral gland preservation. Survival analyses, including overall survival (OS) and recurrence-free survival (RFS), were restricted to the oral cavity SCC cohort (n = 125) to minimize heterogeneity related to tumor site and histopathology.

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Ethics Committee of Gazi University (approval no: 2026-13; date: 13/01/2026). Due to the retrospective design, informed consent was waived in accordance with institutional regulations.

All patients who presented with various symptoms and were diagnosed with head and neck cancer based on imaging and biopsy results were clinically evaluated. The local and/or regional spread of the tumors was assessed through detailed physical examination and necessary radiological imaging methods [computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) etc.]. After diagnosis and preoperative staging, all cases were discussed at a multidisciplinary oncology council meeting. In accordance with the council's decisions, surgical excision was performed in all patients deemed appropriate.

All patients were staged using the American Joint Committee on Cancer staging system (8th edition) based on clinical, radiological, and pathological data obtained during the preoperative and postoperative periods (16).

Based on the pathology data obtained after surgery, patients were classified according to the histopathological data and anatomical locations of head and neck cancer. Level Ib involvement was assessed in all patients (n = 175). Histopathological SMG involvement was assessed in patients with at least one excised SMG specimen (n = 165).

Patients who underwent submandibular gland-sparing surgery (n = 15) were identified based on the following characteristics at the time of their original procedure:

- Patients who were evaluated preoperatively with imaging and clinically as N0,
- Patients who underwent elective neck dissection,
- Patients who underwent bilateral neck dissection with SMG preservation performed on the elective clinically N0 side.

Among these patients, 12 had oral cavity SCC and were included in the survival analyses comparing gland-sparing and non-gland-sparing surgery.

In bilateral neck dissections, SMG preservation was performed only on the elective clinically N0 side; however, contralateral therapeutic SMG excision was performed in 5 of these patients.

The preoperative appearance of the lower lip in a patient with lower lip SCC who underwent submandibular gland-sparing surgery (Figure 1), along with the intraoperative views of the Level Ib dissection before (Figure 2) and after dissection (Figure 3), are presented.

Statistical Analysis

Statistical analyses were performed using SPSS software (IBM SPSS Statistics 27). Descriptive statistics included continuous variables as mean ± standard deviation or median (minimum-maximum) and categorical variables as number and percentage (%). Pearson's chi-square test was used to evaluate relationships between categorical variables in intergroup comparisons. When expected cell counts were <5, Fisher's exact test was used. Some categorical variables were grouped and recoded prior to significance analyses; for example, anatomical tumor subtypes were divided into three main groups: Oral cavity, salivary gland, and others. The Kaplan-Meier method was used for survival analyses, and the log-rank test was used to compare differences between groups. Survival analyses were performed in the oral cavity SCC cohort (n = 125), comparing patients who underwent submandibular gland-sparing surgery (n = 12) with those who underwent non-gland-sparing surgery (n = 113). The clinicopathological characteristics of the gland-sparing and non–gland-sparing groups within the oral cavity SCC cohort were summarized descriptively. No formal between-group hypothesis testing was performed for these characteristics because of the small and markedly unequal group sizes. Follow-up duration was defined as the interval from surgery to death or the last documented clinical contact. Median follow-up was estimated using the reverse Kaplan-Meier method. OS was calculated from the date of surgery to death from any cause, with surviving patients censored at the last documented follow-up. RFS was defined as the interval from the date of surgery to the first documented local, regional, or distant recurrence. Patients who remained recurrence-free were censored at the last documented follow-up, whereas patients who died without a documented recurrence were censored at the date of death.

Figure 1. Preoperative appearance of lower lip squamous cell carcinoma in a patient who underwent submandibular gland-sparing Level Ib neck dissection (gland-sparing case).

RESULTS

The general characteristics of the patients who underwent neck dissection are summarized in Table 1. A total of 175 patients were included in this study. Of the 175 patients, the SMG was available for evaluation in 165; direct gland involvement was identified in 1/165 (0.6%). The predominant histology was SCC (n = 156, 89.1%). The mean age of the patients was 65.8 ± 13.8 years (range, 18-94 years). Fifty-eight percent of the patients were male (n = 102) and 42% were female (n = 73). Histopathological examination after surgery revealed that SCC was the most common type (89.1%). This was followed by mucoepidermoid carcinoma (3.4%), adenoid cystic carcinoma (1.7%), and other types (5.7%) of salivary gland cancer. In terms of anatomical localization, the most common tumor location was the oral cavity (72.0%). This was followed by laryngeal (9.1%), salivary gland (7.4%), neck skin (4.0%), maxillary and paranasal sinus (3.4%), oropharyngeal, nasopharyngeal, and SCC cases of unknown primary origin.

When patients were examined in terms of staging:

• pT4 was the most common stage among the 175 patients, observed in 47 cases (26.9%) in this group. pT4 is followed by pT1 (n = 43, 24.6%), pT2 (n = 42, 24%), pT3 (n = 34, 19.4%), and pTX (n = 9, 5.1%) stages. In patients classified as pTX, the size or degree of invasion of the primary tumor was not determined.

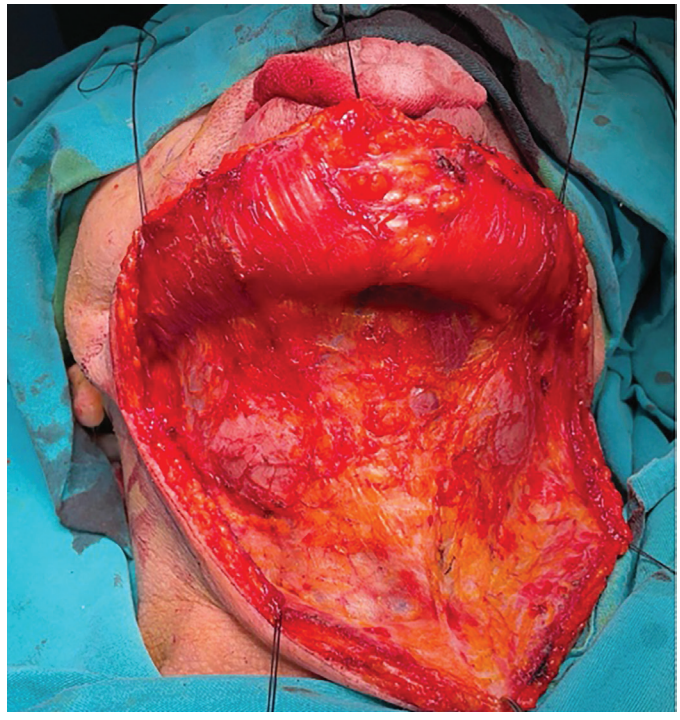

Figure 2. Intraoperative view before Level Ib dissection in the submandibular gland-sparing approach (same patient as Figure 1).

• When evaluated in terms of pN stage, most patients were in the pN0 stage, with this group being the largest subgroup, comprising 98 patients (56%). The distributions of the other nodal stages were as follows: pN3 (n = 31, 17.7%), pN2 (n = 24, 13.7%), pN1 (n = 19, 10.9%), and pNX (n = 3, 1.7%). The pNX stage refers to patients whose regional lymph node status could not be assessed.

Therapeutic neck dissection was performed in 62.3% of the patients, whereas elective neck dissection was performed in 37.7%. The seemingly high frequency of therapeutic neck dissection in our cohort is due to cases with clinically and radiologically suspicious or positive lymph nodes on preoperative imaging (CT/MRI/PET) being classified as therapeutic; nevertheless, some of these cases were ultimately found to be pN0 on final histopathological examination.

When the postoperative pathology reports of the patients were examined, Level Ib involvement was detected in 34 patients (19.4%), whereas no involvement was detected in 141 patients (80.6%). SMG involvement was observed in only one patient (0.6%) who had undergone surgery for lower lip SCC. In this patient, metastatic disease was observed in the ipsilateral level Ib, with involvement of the SMG likely resulting from direct extension of the metastatic lymph node. Submandibular gland-sparing surgery was performed in 15 patients (8.6%). The pathological evaluation demonstrating SMG invasion is presented in Figure 4.

The distribution of patients with level Ib lesions according to anatomical localization is shown in Table 2. Level Ib involvement showed significant differences according to the anatomical localization of the tumor (p = 0.005). The rate of level Ib involvement is higher, especially in tumors originating from the salivary glands and oral cavity.

Table 3 presents level Ib involvement according to oral cavity subsites. Due to limited sample sizes in certain subsites, these areas were combined for analysis. No statistically significant differences in level Ib involvement were detected among the oral cavity subsites (p = 0.112). However, higher rates of level Ib involvement were observed in the buccal plus retromolar subgroup (37.5%) and the combined alveolar/mandible, hard palate, and floor of mouth subgroup (30.0%), although these differences did not achieve statistical significance.

Table 4 presents the main clinicopathological characteristics of the oral cavity SCC cohort according to SMG preservation status. The mean age was 62.2 ± 11.0 years in the gland-sparing group and 67.0 ± 13.1 years in the non-gland-sparing group. Tongue tumors were the most frequent subsite in both groups, accounting for 33.3% and 39.8% of cases, respectively. In the gland-sparing group, 7 of 12 patients (58.3%) had pT1-pT2 disease, 4 (33.3%) had pT4 disease, and 9 (75.0%) were pN0 on final pathological staging. The corresponding proportions in the non-gland-sparing group were 61.1%, 18.6%, and 62.8%, respectively. These characteristics are presented descriptively to illustrate the composition of the two clinically selected groups.

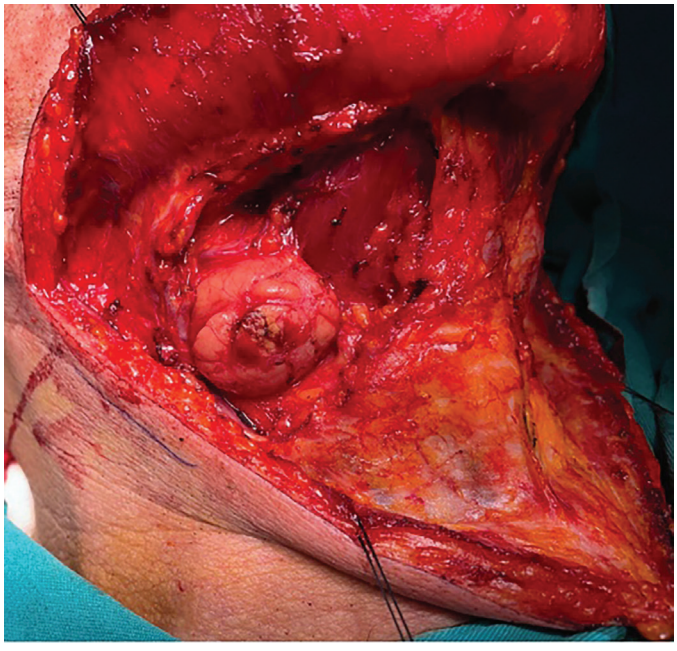

Figure 3. Intraoperative view after completion of Level Ib dissection with preservation of the submandibular gland (same patient as Figure 1).

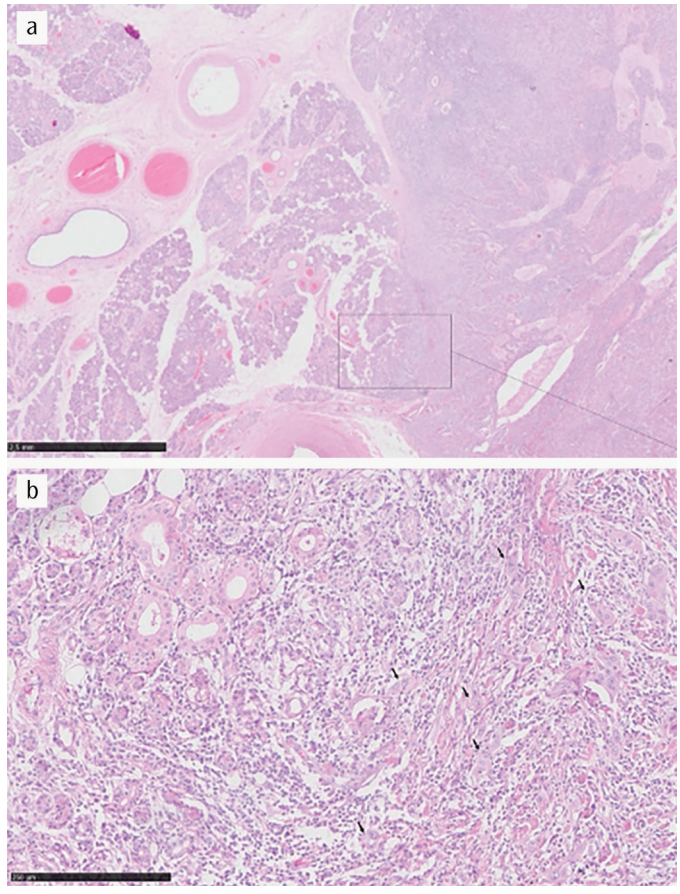

Figure 4. Pathological specimen image demonstrating submandibular gland involvement (non-gland-sparing case).

a) Low-power histological section demonstrating squamous cell carcinoma infiltrating the submandibular gland parenchyma (H&E, x13). The boxed area indicates the region shown at a higher magnification. b) High-power view of the highlighted area showing infiltrating nests and cords of squamous cell carcinoma (arrow) within the submandibular gland stroma (H&E, x140).

H&E: Hematoxylin and eosin.

Table 1. General characteristics of patients who underwent neck dissection.

Variable (n=175)	n	Percentage (%)	Mean ± SD/min-max
Age			**65.8 ± 13.8**
			18-94
Gender			
Male	102	58	
Female	73	42	
Histopathologic subtype of tumor			
SCC	156	89.1	
Mucoepidermoid carcinoma	6	3.4	
Adenoid cystic carcinoma	3	1.7	
Others	10	5.7	
Anatomical subtype of head and neck cancer			
Oral cavity	126	72	
Oropharynx	3	1.7	
Larynx	16	9.1	
Salivary gland	13	7.4	
Nasopharynx	2	1.1	
Maxilla and paranasal sinus	6	3.4	
Unknown primary SCC	2	1.1	
Skin cancer of the head and neck	7	4.0	
Type of neck dissection			
Elective	66	37.7	
Therapeutic	109	62.3	
Level Ib involvement			
Present	34	19.4	
Absent	141	80.6	
Submandibular gland involvement (n = 165)			
Present	1	0.6	
Absent	164	99.4	
Submandibular gland-sparing surgery			
Present	15	8.6	
Absent	160	91.4	
pT stage			
pT1	43	24.6	
pT2	42	24	
pT3	34	19.4	
pT4	47	26.9	
pTX	9	5.1	
pN stage			
pN0	98	56	
pN1	19	10.9	
pN2	24	13.7	
pN3	31	17.7	
pNX	3	1.7	

SCC: Squamous cell carcinoma.

Table 2. Distribution of patients with level Ib lesions according to anatomical localization.

Tumor location	Level Ib involvement present (n, %)	Level Ib involvement absent (n, %)	Total (n, %)	Pearson chi-square (χ^2)	df	p-value
Oral cavity (group 1)	26 (20.6%)	100 (79.4%)	126 (72.0%)			
Salivary gland (group 2)	6 (46.2%)	7 (53.8%)	13 (7.4%)			
Other sites (group 3)	2 (5.6%)	34 (94.4%)	36 (20.6%)			
Total	34 (19.4%)	141 (80.6%)	175 (100%)	10.475	2	**0.005**

Table 3. Level Ib involvement according to oral cavity subsites.

Oral cavity subsite	Level Ib involvement present n (%)	Level Ib involvement absent n (%)	Total n (%)	Pearson chi-square (χ^2)	df	p-value
Lip	8 (19.5)	33 (80.5)	41 (32.5%)			
Tongue	6 (12.2)	43 (87.8)	49 (38.9%)			
Alveolar/mandible + hard palate + floor of mouth	6 (30.0)	14 (70.0)	20 (15.9%)			
Buccal + retromolar	6 (37.5)	10 (62.5)	16 (12.7%)			
Total	**26 (20.6)**	**100 (79.4)**	**126 (100%)**	**5.988**	**3**	**0.112**

Table 4. Clinicopathological characteristics of the oral cavity SCC cohort according to submandibular gland preservation status.

Characteristic	Non-gland-sparing (n = 113)	Gland-sparing (n = 12)	Total (n = 125)
Age, years, mean ± SD	67.0 ± 13.1	62.2 ± 11.0	66.5 ± 13.0
Sex			
Male	58 (51.3)	8 (66.7)	66 (52.8)
Female	55 (48.7)	4 (33.3)	59 (47.2)
Tumor subsite			
Lip	38 (33.6)	3 (25.0)	41 (32.8)
Tongue	45 (39.8)	4 (33.3)	49 (39.2)
Alveolar/mandible + hard palate + floor of mouth	17 (15.0)	2 (16.7)	19 (15.2)
Buccal + retromolar	13 (11.5)	3 (25.0)	16 (12.8)
pT stage			
pT1	38 (33.6)	3 (25.0)	41 (32.8)
pT2	31 (27.4)	4 (33.3)	35 (28.0)
pT3	20 (17.7)	1 (8.3)	21 (16.8)
pT4	21 (18.6)	4 (33.3)	25 (20.0)
pTX	3 (2.7)	0 (0.0)	3 (2.4)
pN stage			
pN0	71 (62.8)	9 (75.0)	80 (64.0)
pN1	13 (11.5)	1 (8.3)	14 (11.2)
pN2	12 (10.6)	1 (8.3)	13 (10.4)
pN3	17 (15.0)	1 (8.3)	18 (14.4)

SCC: Squamous cell carcinoma.

Tables 5 and 6 present the effects of submandibular gland-sparing surgery on OS and RFS in the oral cavity SCC cohort. Because gland-sparing procedures were performed in selected oral cavity SCC cases, survival analyses were restricted to this subgroup to minimize heterogeneity related to tumor site and histopathology.

The median follow-up duration for the oral cavity SCC survival cohort was 74.0 months (observed range, 1-116 months), as estimated using the reverse Kaplan-Meier method.

The effect of submandibular gland-sparing surgery on OS is shown in Table 5. In the oral cavity SCC cohort, 12 patients underwent gland-sparing surgery, whereas 113 patients underwent non-gland-sparing surgery. Kaplan-Meier log-rank analysis showed no statistically significant difference in OS between the gland-sparing and non-gland-sparing groups (mean OS: 88.6 ± 7.1 vs. 84.8 ± 4.4 months, respectively; log-rank χ^2 = 1.128, p = 0.288).

The effect of submandibular gland-sparing surgery on RFS is shown in Table 6. No statistically significant difference in RFS was observed between the gland-sparing and non-gland-sparing groups (mean RFS: 74.2 ± 10.9 vs. 94.4 ± 4.2 months, respectively; log-rank χ^2 = 0.006, p = 0.936).

DISCUSSION

A review of the literature revealed that SMG involvement in head and neck cancers occurs at a notably low rate. In the absence of risk factors, submandibular gland-preserving surgery can be considered within the framework of oncologically safe. In our study, the detection of SMG involvement in only one patient out of a sample of 165 supports the existing literature.

Spiegel et al. (13) studied the involvement of SMG in neck dissections (unilateral or bilateral) performed for head and neck SCC in 169 subjects. 196 SMG were resected. Glandular invasion was observed in nine SMG (4.6%), glandular invasion was observed (3 cases; extension from level Ib lymph node metastasis, and six cases showed direct extensions from the primary tumor). They suggested that transplantation or replantation of the contralateral SMG may be a safe oncological procedure (13).

Subramaniam et al. (17) studied the determinants of level Ib involvement in 585 neck dissections from an OSCC cohort. In 226 N+ cases, 21 (9.3%) had direct tumor invasion of the SMG. The risk factors for SMG involvement were tumor diameter >4 cm and DOI >10 mm. Based on their findings, they proposed a submandibular gland-sparing neck dissection for early stage OSCC. In a study by Kalloli et al. (18), SMG involvement was observed in five (1.6%) of 281 patients with OSCC who underwent neck dissection. SMG involvement was from level Ib metastasis in three cases (0.9%) and from direct extension of the primary tumor in two cases (0.6%). Pasha et al. (19) studied involvement of SMG in a case series (149 patients) of oral cavity SCC. The SMG was involved in seven cases (4.7%), all of whom had level Ib metastasis. In addition, two of them had direct extension from the primary origin (floor of mouth). Iocca et al. (20) designed a retrospective multicenter study of SMG involvement in 642 patients with oral cavity SCC. SMG involvement was observed in 12/642 (1.9%) and 12/852 (1.4%) per patient and gland, respectively. pN0 cases were associated with reduced SMG involvement. In addition, they performed a meta-analysis of SMG involvement in 4458 cases and 5037 glands. The involvement rates were 1.8% [99% confidence interval (CI) 1.1-2.7] and 1.6% (99% CI 1.0-2.4), respectively.

Gu et al. (21) performed a prospective non-randomized cohort study examining the impact of SMG preservation in early stage buccal SCC. A total of 137 patients were included in this study. A total of 31 patients underwent SMG-sparing neck dissection. The 5-year locoregional control rates in the SMG-sparing and SMG-excision groups were 74% and 75%, respectively (p = 0.970). In addition, the 5-year DSS rates were 74% and 69%, respectively. Likewise, Du et al. (22) retrospectively studied the outcomes of SMG preservation and excision in neck dissection for cT1-2N0 floor of mouth SCC. The 5-year LCR rates for SMG preservation and excision groups were found to be 84% and 73%, respectively.

In another study on this subject, a retrospective study of 317 cases conducted by Rahman et al. (23) reported a SMG involvement rate of 3.8%. The study showed that the most important risk factors for SMG involvement were level Ib lymph node involvement [odds ratio (OR), 13.6], extranodal spread (OR, 67), and perineural invasion (OR, 5). It has been emphasized that in early stage buccal mucosal cancers at the cT1-T2N0 stage, preservation of the SMG may be possible without compromising oncological safety in the

Table 5. Effect of submandibular gland-sparing surgery on overall survival in the oral cavity SCC cohort.

Submandibular gland-sparing surgery	Total patients	Number of events	Mean OS (months) ± SE	Log-Rank χ^2	df	p-value
No	113	34	84.8 ± 4.4			
Yes	12	1	88.6 ± 7.1			
Total	125	35	86.1 ± 4.2	1.128	1	0.288

SCC: Squamous cell carcinoma, OS: Overall survival, SE: Standard error.

Table 6. Effect of submandibular gland-sparing surgery on recurrence-free survival in the oral cavity SCC cohort.

Submandibular gland-sparing surgery	Total patients (N)	Number of events (N)	Mean RFS (months) ± SE	Log-Rank χ^2	df	p-value
No	113	21	94.4 ± 4.2			
Yes	12	2	74.2 ± 10.9			
Total	125	23	94.2 ± 4.1	0.006	1	0.936

SCC: Squamous cell carcinoma, SE: Standard error, RFS: Recurrence-free survival.

absence of these high-risk factors. Finally, Jakhetiya et al. (24) reported an SMG involvement rate of 1.32% in their retrospective study of 303 patients who underwent neck dissection due to oral cavity SCC. The study noted that submandibular invasion occurred either through direct infiltration of the primary tumor or through extracapsular spread from adjacent metastatic lymph nodes, and that isolated intraglandular metastasis was not observed.

When the data from our study were examined, the pathological data of 165 patients who underwent surgery for head and neck cancer revealed that SMG involvement was detected in only one patient (0.6%), which is a very low rate of occurrence. This finding is similar to the 1.32% rate reported by Jakhetiya et al. (24) and the rates of 1.9% per patient and 1.4% per gland reported in Iocca et al.'s (20) multicenter retrospective study. In the literature, tumor diameter >4 cm, depth of invasion (DOI) >10 mm, level Ib lymph node involvement, extranodal spread, and perineural invasion have been identified as the most important risk factors for SMG involvement. Since SMG involvement was detected in only one patient in our study, risk factor analysis could not be performed, which is one of the main limitations of our study.

In the present study, survival analyses were restricted to the oral cavity SCC cohort to minimize heterogeneity related to tumor site and histopathology, as gland-sparing procedures were performed in selected oral cavity SCC cases. No statistically significant difference was observed in OS or RFS between the gland-sparing and non-gland-sparing groups ($p = 0.288$ and $p = 0.936$, respectively). These findings are in line with the studies by Gu et al. (21) and Du et al. (22), which reported comparable oncological outcomes between SMG-sparing and SMG-excision approaches in selected oral cavity SCC cohorts. Nevertheless, the survival analysis included only 12 patients in the gland-sparing group, with one OS event and two RFS events. These low event counts limited the ability to detect between-group differences; therefore, the absence of statistically significant differences should not be interpreted as evidence of equivalence between the two surgical approaches. Together with the retrospective, non-randomized design and potential selection bias, these survival findings should be considered exploratory.

Study Limitations

The main limitations of this study include its retrospective design, the inability to perform risk factor analysis due to the very low number of SMG involvement events, and the limited sample size of the gland-sparing subgroup. In addition, gland-sparing surgery was performed only in selected clinically N0 oral cavity SCC cases, which may have introduced selection bias. Therefore, the survival analysis should be regarded as exploratory rather than confirmatory.

CONCLUSION

In the present study, SMG involvement was identified in 0.6% of patients who underwent neck dissection for head and neck malignancy. In the oral cavity SCC cohort, no statistically significant difference was found in OS or RFS between patients who underwent submandibular gland-sparing surgery and those who did not. Although the limited number of SMG involvement events precluded a comprehensive risk factor analysis, our findings support the existing literature suggesting that SMG preservation may be a feasible approach in carefully selected low-risk oral cavity SCC patients without clinical or radiological evidence of level Ib nodal involvement.

Ethics

Ethics Committee Approval: The study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Ethics Committee of Gazi University (approval no: 2026-13; date: 13/01/2026).

Informed Consent: The requirement for informed consent was waived due to the retrospective design; however, written informed consent for publication of clinical and intraoperative images was obtained from the patient.

Acknowledgments

We acknowledge the use of ChatGPT (OpenAI) for language editing and translation support.

Footnotes

Authorship Contributions

Surgical and Medical Practices: E.A., B.Ö., U.A., M.Y., M.D., Concept: M.A., Design: M.A., E.A., Data Collection or Processing: M.E.Z., E.A., Analysis or Interpretation: M.E.Z., B.Ö., U.A., M.Y., M.D., Literature Search: M.A., M.E.Z., Writing: M.A., B.Ö., U.A., M.Y., M.D.

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

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

REFERENCES

1. Crile G. Landmark article Dec 1, 1906: excision of cancer of the head and neck. With special reference to the plan of dissection based on one hundred and thirty-two operations. By George Crile. JAMA. 1987; 258: 3286-93.
2. Myers EN, Gastman BR. Neck dissection: an operation in evolution: Hayes Martin lecture. Arch Otolaryngol Head Neck Surg. 2003; 129: 14-25.
3. Harish K. Neck dissections: radical to conservative. World J Surg Oncol. 2005; 3: 21.
4. Shah JP. Patterns of cervical lymph node metastasis from squamous carcinomas of the upper aerodigestive tract. Am J Surg. 1990; 160: 405-9.
5. Cadoni G, Giraldi L, Petrelli L, Pandolfini M, Giuliani M, Paludetti G, et al. Prognostic factors in head and neck cancer: a 10-year retrospective analysis in a single-institution in Italy. Acta Otorhinolaryngol Ital. 2017; 37: 458-66.
6. Pisani P, Airoldi M, Allais A, Aluffi Valletti P, Battista M, Benazzo M, et al. Metastatic disease in head & neck oncology. Acta Otorhinolaryngol Ital. 2020; 40(Suppl 1): S1-86.
7. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. Nat Rev Dis Primers. 2020; 6: 92.
8. Weiss MH, Harrison LB, Isaacs RS. Use of decision analysis in planning a management strategy for the stage N0 neck. Arch Otolaryngol Head Neck Surg. 1994; 120: 699-702.

9. Ferlito A, Robbins KT, Silver CE, Hasegawa Y, Rinaldo A. Classification of neck dissections: an evolving system. Auris Nasus Larynx. 2009; 36: 127-34.
10. Byers RM, Wolf PF, Ballantyne AJ. Rationale for elective modified neck dissection. Head Neck Surg. 1988; 10: 160-7.
11. Kowalski LP, Sanabria A. Elective neck dissection in oral carcinoma: a critical review of the evidence. Acta Otorhinolaryngol Ital. 2007; 27: 113-7.
12. Robbins KT, Clayman G, Levine PA, Medina J, Sessions R, Shaha A, et al. Neck dissection classification update: revisions proposed by the American Head and Neck Society and the American Academy of Otolaryngology-Head and Neck Surgery. Arch Otolaryngol Head Neck Surg. 2002; 128: 751-8.
13. Spiegel JH, Brys AK, Bhakti A, Singer MI. Metastasis to the submandibular gland in head and neck carcinomas. Head Neck. 2004; 26: 1064-8.
14. Jaguar GC, Lima EN, Kowalski LP, Pellizon AC, Carvalho AL, Alves FA. Impact of submandibular gland excision on salivary gland function in head and neck cancer patients. Oral Oncol. 2010; 46: 349-54.
15. Chiesa Estomba C, Valldeperes-Vilanova A, Gonzalez-Garcia J, Larruscain-Sarasola E, Sistiaga-Suarez J, Altuna-Mariezcurrena X. Neurological complications and quality of life after submandibular gland resection. A prospective, non-randomized, single-centre study. Otolaryngol Pol. 2019; 73: 32-7.
16. Amin MB, Edge SB, Greene FL, Byrd DR, Brookland RK, Washington MK, et al. AJCC cancer staging manual: Springer; 2017.
17. Subramaniam N, Balasubramanian D, Reddy R, Rathod P, Murthy S, Vidhyadharan S, et al. Determinants of level Ib involvement in oral squamous cell carcinoma and implications for submandibular gland-sparing neck dissection. Int J Oral Maxillofac Surg. 2018; 47: 1507-10.
18. Kalloli M, Patil RS, Singh R, Prabhu N, Issrani R. Submandibular gland extirpation during neck dissection, is it truly justified? Indian J Cancer. 2022; 59: 591-6.
19. Pasha HA, Dhanani R, Ghaloo SK, Ghias K, Khan MJ. Level I nodal positivity as a factor for involvement of the submandibular gland in oral cavity carcinoma: a case series report. Int Arch Otorhinolaryngol. 2021; 25: e279-83.
20. Iocca O, Copelli C, Garzino-Demo P, Ramieri G, Rubattino S, Sedran L, et al. Submandibular gland involvement in oral cavity squamous cell carcinoma: a retrospective multicenter study. Eur Arch Otorhinolaryngol. 2023; 280: 4205-14.
21. Gu B, Fang Q, Wu Y, Du W, Zhang X, Chen D. Impact of submandibular gland preservation in neck management of early-stage buccal squamous cell carcinoma on locoregional control and disease-specific survival. BMC Cancer. 2020; 20: 1034.
22. Du W, Fang Q, Liu S, Chen D, Luo R, Zhang X. Feasibility of submandibular gland preservation in cT1-2N0 squamous cell carcinoma in the floor of the mouth. Front Oncol. 2020; 10: 579.
23. Rahman T, Khandelwal K, Bannoth S, Kumar Das A, Das R, Das K, et al. Submandibular gland in neck dissection: necessary fatality or unfortunate prey? Indian J Otolaryngol Head Neck Surg. 2023;75: 789-93.
24. Jakhetiya A, Kaul P, Pandey A, Patel T, Kumar Meena J, Pal Singh M, et al. Distribution and determinants of submandibular gland involvement in oral cavity squamous cell carcinoma. Oral Oncol. 2021; 118: 105316.

Original Investigation | Özgün Araştırma

Gazi Med J 2026;37(4):533-537

DOI: http://dx.doi.org/10.12996/gmj.2026.4752

Long-term Outcomes of Kidney Transplant Recipients with Delayed Graft Function: A Single-center Experience

Gecikmiş Greft Fonksiyonu Gelişen Böbrek Nakli Alıcılarının Uzun Dönem Sonuçları: Tek Merkez Deneyimi

Selim Keçeoğlu[1,2], Mesut Yavaş[1], Ramazan Kozan[1], Aydın Dalgıç[1], Hakan Sözen[1,2]

[1]Department of General Surgery, Gazi University, Faculty of Medicine, Ankara, Türkiye
[2]Department of Organ Transplantation Center, Gazi University, Faculty of Medicine, Ankara, Türkiye

ABSTRACT

Objective: Delayed graft function (DGF), defined as the need for dialysis within the first week after transplantation, is a common early complication after deceased donor kidney transplantation. Although it is associated with adverse short-term outcomes, its long-term impact remains uncertain. We evaluated long-term patient and graft outcomes among deceased donor kidney transplant recipients who developed DGF.

Methods: This retrospective study identified 92 adult recipients of first deceased donor kidney transplants, of whom 79 developed DGF and constituted the primary analytic cohort. DGF was defined as dialysis requirement within 7 days post-transplant. Rejection, infections, graft loss, mortality, and long-term survival were analyzed among the 79 recipients who developed DGF.

Results: DGF occurred in 79 of 92 recipients (85.9%). Median follow-up was 113 months. Among the 79 recipients with DGF, six acute rejection episodes were observed, including two antibody-mediated cases; all rejection episodes received treatment according to the type of rejection. Among the 79 recipients with DGF, five viral infections were documented (BK virus n = 3, cytomegalovirus n = 1, parvovirus n = 1). During follow-up, 22 graft losses and 13 deaths occurred. Patient survival at 1, 3, 5, and 10 years was 100%, 97%, 93%, and 89%, respectively. Death-censored graft survival rates were 100%, 97%, 93%, and 86%, respectively.

ÖZ

Amaç: Transplantasyon sonrası ilk hafta içinde diyaliz gereksinimi olarak tanımlanan gecikmiş greft fonksiyonu (*delayed graft function*, DGF), kadavra vericiden böbrek nakli sonrasında sık görülen erken dönem komplikasyonlardan biridir. Kısa dönem olumsuz sonuçlarla ilişkili olmasına rağmen, uzun dönem etkisi belirsizliğini korumaktadır. Bu çalışmada, DGF gelişen kadavra vericiden böbrek nakli alıcılarında uzun dönem hasta ve greft sonuçlarının değerlendirilmesi amaçlandı.

Yöntemler: Bu retrospektif çalışmada, ilk kez kadavra vericiden böbrek nakli yapılan 92 erişkin alıcı değerlendirildi. Bunların 79'unda DGF gelişti ve bu hastalar temel analiz kohortunu oluşturdu. DGF, transplantasyon sonrası 7 gün içinde diyaliz gereksinimi olarak tanımlandı. DGF gelişen 79 alıcıda rejeksiyon, enfeksiyonlar, greft kaybı, mortalite ve uzun dönem sağkalım analiz edildi.

Bulgular: DGF, 92 alıcının 79'unda (%85,9) gelişti. Ortanca takip süresi 113 aydı. DGF gelişen 79 alıcıda, ikisi antikor aracılı olmak üzere toplam altı akut rejeksiyon atağı gözlendi; tüm rejeksiyon atakları, rejeksiyon tipine uygun olarak tedavi edildi. DGF gelişen 79 alıcı arasında beş viral enfeksiyon saptandı (BK virüsü, n = 3; sitomegalovirüs, n = 1; parvovirüs, n = 1). Takip süresince 22 greft kaybı ve 13 ölüm meydana geldi. Hasta sağkalımı 1., 3., 5. ve 10. yıllarda sırasıyla %100, %97, %93 ve %89 olarak bulundu. Ölüm sansürlü greft sağkalım oranları ise sırasıyla %100, %97, %93 ve %86 idi.

Cite this article as: Keçeoğlu S, Yavaş M, Kozan R, Dalgıç A, Sözen H. Long-term outcomes of kidney transplant recipients with delayed graft function: a single-center experience. Gazi Med J. 2026;37(4):533-537

Address for Correspondence/Yazışma Adresi: Selim Keçeoğlu, Department of General Surgery; Department of Organ Transplantation Center, Gazi University, Faculty of Medicine, Ankara, Türkiye
E-mail / E-posta: selimkeceoglu@gmail.com
ORCID ID: orcid.org/0000-0002-9985-3972

Received/Geliş Tarihi: 26.06.2026
Accepted/Kabul Tarihi: 08.09.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

Conclusion: In this cohort of deceased donor kidney transplant recipients who developed DGF, favorable long-term patient and graft outcomes were observed. These findings suggest that favorable long-term outcomes may be achieved among recipients who develop DGF when appropriate post-transplant management is provided. The occurrence of viral infections highlights the need for close monitoring and individualized immunosuppressive strategies.

Keywords: Delayed graft function, kidney transplantation, deceased donor kidney transplantation, graft survival, patient survival, acute rejection, long-term outcomes, immunosuppressive therapy

ÖZ

Sonuç: DGF gelişen kadavra vericiden böbrek nakli alıcılarından oluşan bu kohortta, uzun dönem hasta ve greft sonuçlarının olumlu olduğu gözlendi. Bu bulgular, uygun nakil sonrası yönetim sağlandığında DGF gelişen alıcılarda olumlu uzun dönem sonuçlara ulaşılabileceğini düşündürmektedir. Viral enfeksiyonların görülmesi, yakın izlemin ve bireyselleştirilmiş immünosüpresif tedavi stratejilerinin gerekliliğini ortaya koymaktadır.

Anahtar Sözcükler: Gecikmiş greft fonksiyonu, böbrek transplantasyonu, kadavra vericiden böbrek nakli, greft sağkalımı, hasta sağkalımı, akut rejeksiyon, uzun dönem sonuçlar, immünosüpresif tedavi

INTRODUCTION

Kidney transplantation is the preferred treatment for patients with end-stage kidney disease, offering improved patient survival, enhanced health-related quality of life, and reduced long-term healthcare expenditures compared with maintenance dialysis (1,2). Although substantial progress has been achieved in operative procedures, graft preservation methods, and immunosuppressive management, early post-transplant events remain an important determinant of clinical outcomes, especially among deceased donor kidney transplant recipients. Among these, delayed graft function (DGF) remains a common and clinically relevant complication (1).

The most widely accepted definition of DGF is the requirement for dialysis during the first post-transplant week. It is widely considered a manifestation of ischemia-reperfusion injury leading to acute tubular dysfunction and delayed recovery of graft function (3). The incidence of DGF varies according to multiple donor-related, recipient-related, and transplant-related factors, as well as preservation and ischemia times, but it is consistently higher in deceased donor transplantation, particularly in expanded-criteria and donation after circulatory death donors (1,4,5).

The impact of DGF extends beyond the initial post-transplant stage. DGF has been linked to extended hospital stays, greater healthcare resource consumption, an elevated risk of acute rejection, and suboptimal early graft function (6). Nevertheless, the long-term consequences of DGF remain a matter of ongoing debate. While some investigators have observed poorer graft survival among recipients who develop DGF, others have reported satisfactory long-term outcomes when modern immunosuppressive protocols and optimized post-transplant follow-up are applied (6,7). Emerging evidence also indicates that the severity and duration of DGF may be more relevant than its presence alone (8).

These inconsistencies likely reflect differences in donor quality, recipient risk profiles, ischemia exposure, and center-specific practices. Therefore, contemporary data with extended follow-up are needed to better define the long-term implications of DGF.

In this study, we evaluated long-term patient and graft outcomes among deceased donor kidney transplant recipients who developed DGF, with a focus on rejection, infectious complications, graft loss, and long-term survival.

MATERIALS AND METHODS

Study Design and Setting

This retrospective, single-center cohort study was conducted at the Gazi University Transplantation Center, Ankara, Türkiye. All consecutive deceased donor kidney transplantations performed between January 2006 and December 2019 were screened for eligibility. Clinical and demographic data were obtained from the institutional transplant registry, electronic medical records, and archived hospital charts.

Study Population

A total of 92 deceased donor kidney transplant recipients were identified during the study period. Patients undergoing first deceased donor kidney transplantation with available follow-up data were eligible. Patients with re-transplantation, multi-visceral transplantation, or incomplete clinical data were excluded (Figure 1). DGF developed in 79 of the 92 recipients, whereas 13 recipients did not develop DGF. Because the primary objective was to evaluate long-term patient and graft outcomes among recipients who developed DGF, the primary outcome analysis was restricted to the 79 recipients with DGF.

Surgical Technique

Kidney grafts were preferentially implanted in the right iliac fossa; the left iliac fossa or an intraabdominal approach was used when clinically indicated. Vascular reconstruction was performed using

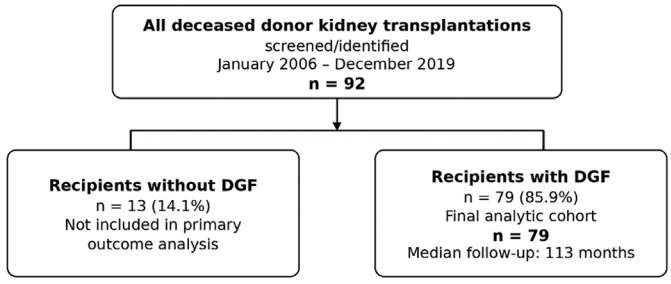

Figure 1. Flowchart of patient selection and formation of the primary analytic cohort.

DGF: Delayed graft function.

standard end-to-side anastomosis techniques, including Carrel patch implantation when appropriate. Ureterovesical anastomosis was carried out using a modified Lich-Gregoire technique. When indicated, the corner-saving modification described by Mehmet Haberal and colleagues was applied (9). Double-J ureteral stents were routinely used and removed postoperatively according to institutional practice. The surgical approach has been described previously (10).

Definition of DGF

DGF was considered present when dialysis was required during the first post-transplant week (11). The decision to initiate dialysis was based on clinical and biochemical parameters, including hyperkalemia, metabolic acidosis, fluid overload, or insufficient early graft function. The same DGF definition and dialysis initiation criteria were applied consistently throughout the study period (2006-2019).

Immunosuppressive Protocol

The institutional immunosuppressive approach has been described previously (10). Briefly, all recipients received a calcineurin inhibitor-based triple maintenance regimen consisting of tacrolimus or cyclosporine, mycophenolate mofetil, and corticosteroids. Induction therapy was administered according to immunological risk and institutional protocols. Interleukin-2 receptor antagonists were used in selected recipients, whereas antithymocyte globulin (ATG) was preferred in patients with high immunological risk or DGF. In recipients with DGF, calcineurin inhibitor initiation could be delayed, and ATG was continued until graft recovery, defined clinically by improving urine output and/or a substantial decline in serum creatinine. The immunosuppressive protocol remained unchanged throughout the study period (2006-2019).

Post-transplant Monitoring and Infection Prophylaxis

Recipients were followed with scheduled outpatient visits, laboratory assessments, serum creatinine monitoring, therapeutic drug level measurements, and surveillance for infectious complications. Standard perioperative antibiotic prophylaxis was administered. Trimethoprim-sulfamethoxazole was used for Pneumocystis jirovecii prophylaxis, and cytomegalovirus (CMV) prophylaxis was provided according to donor-recipient serostatus and institutional protocols.

Study Outcomes

The prespecified outcomes were patient survival, death-censored graft survival, acute rejection episodes, viral and opportunistic infections, and causes of graft loss. Graft loss was defined as permanent loss of graft function requiring return to dialysis or retransplantation. Acute rejection was diagnosed based on clinical and/or histopathological findings according to Banff criteria when biopsy material was available (12). T-cell-mediated rejection was treated with pulse corticosteroids, whereas antibody-mediated rejection was managed with plasmapheresis, intravenous immunoglobulin, and/or lymphocyte-depleting therapy.

Data Collection

The following variables were recorded: recipient sex, donor age, cold ischemia time, number of human leukocyte antigen (HLA) mismatches, duration of ATG therapy, DGF status, acute rejection, infections, graft loss, cause of graft failure, cause of death, and follow-up duration. Variables known to be associated with DGF and post-transplant outcomes were considered during dataset construction (1-3,5-8).

Ethical Approval

This retrospective study was performed in accordance with the ethical principles of the Declaration of Helsinki and institutional requirements. The study protocol received approval from the Gazi University Ethics Commission before data extraction and analysis (approval no: 05.02.2026-E.1456672, date: 05.02.2026). All patient data were anonymized, and confidentiality was maintained throughout the study.

Statistical Analysis

The primary analysis included the 79 recipients who developed DGF. All outcome analyses, including patient survival and death-censored graft survival, were performed in this primary analytic cohort. Descriptive statistics for continuous variables were reported as mean ± standard deviation or median (minimum-maximum), as appropriate. Categorical data were described using absolute numbers and corresponding percentages. Survival outcomes, including patient survival and death-censored graft survival, were evaluated by Kaplan-Meier survival curves. Kaplan-Meier estimates were used to describe patient survival and death-censored graft survival at 1, 3, 5, and 10 years among the 79 recipients in the primary analytic cohort. Given the retrospective design and limited sample size, analyses were primarily descriptive, and findings were interpreted within the context of an observational study.

RESULTS

Baseline Characteristics

A total of 92 deceased donor kidney transplant recipients were identified during the study period. DGF occurred in 79 recipients (85.9%), whereas 13 recipients did not develop DGF. Among the 79 recipients with DGF, 40 were female and 39 were male. The mean donor age was 41 ± 19 years, mean cold ischemia time was 21 ± 6.1 hours, and mean number of HLA mismatches was 3 ± 2.1. The mean duration of rabbit ATG therapy was 9.4 ± 4 days.

Follow-up Duration

The primary analytic cohort was followed for a mean duration of 98 ± 59 months (range, 6-173 months), with a median follow-up of 113 months.

Rejection Outcomes

Among the 79 recipients with DGF, 6 acute rejection episodes (7.6%) were documented, including 2 antibody-mediated rejections. All rejection episodes received treatment according to the type of rejection.

Infectious Complications

Among the 79 recipients with DGF, 5 viral infections were observed: BK virus (BKV) in 3, CMV in 1, and parvovirus in 1. During the follow-

up period, no instances of post-transplant lymphoproliferative disorder or newly developed malignancy were recorded.

Graft Loss

Among the 79 recipients with DGF, 22 graft losses occurred during follow-up. The leading cause was chronic allograft nephropathy (n = 12), followed by BKV nephropathy (n = 3), hemolytic uremic syndrome (n = 2), sepsis-related graft failure (n = 2), infection originating from the native kidneys (n = 1), CMV-associated injury (n=1), and unexplained causes (n = 1).

Patient Mortality

Among the 79 recipients with DGF, 13 recipients (16.4%) died during follow-up. The most common cause of death was sepsis (n = 7), followed by aortic dissection (n = 1) and traffic accident (n = 1). In the remaining cases, the cause of death could not be determined.

Long-term Survival Outcomes

The estimated patient survival rates at 1, 3, 5, and 10 years were 100%, 97%, 93%, and 89%, respectively. Corresponding death-censored graft survival rates were 100%, 97%, 93%, and 86% at the same time points.

DISCUSSION

In this single-center cohort of recipients who developed DGF, long-term patient and graft survival were favorable. These findings indicate that favorable long-term outcomes can be observed among recipients who develop DGF, although the absence of a formal non-DGF comparator analysis precludes determining whether DGF itself influences long-term outcomes (1,6). Our findings should be interpreted in the context of the high prevalence of DGF in this cohort. The exceptionally high DGF incidence observed in our cohort exceeds rates reported in most contemporary series (20-50%). This difference may be attributable to the exclusively deceased-donor population, prolonged cold ischemia times, and center-specific donor characteristics. Importantly, the same operational definition of DGF and clinical criteria for dialysis initiation were applied throughout the study period, making a change in diagnostic practice an unlikely explanation for the high observed incidence.

A likely contributing factor is the relatively young donor age in our cohort. Younger donor kidneys may better tolerate ischemia-reperfusion injury and recover function despite early dysfunction, which may have mitigated the long-term impact of DGF in this series (2,3).

DGF has been linked to higher early morbidity, including acute rejection and delayed recovery. In our cohort, rejection episodes were infrequent and controllable. The use of ATG-based induction in selected patients with DGF may have contributed to the low acute rejection rate observed in our cohort, potentially reflecting the relatively prolonged ATG exposure during delayed graft recovery (7,13).

Previous studies suggest that the duration and severity of DGF may be more relevant determinants of long-term outcomes than its mere presence. In our cohort, chronic allograft nephropathy accounted for the majority of graft losses, suggesting that long-term graft failure may also reflect ongoing chronic allograft injury rather than the initial episode of DGF alone.

Viral infections, particularly BKV, were observed during follow-up. Recipients with DGF may be exposed to more intensive immunosuppression, which can increase susceptibility to opportunistic infections. This highlights the importance of close monitoring and timely adjustment of immunosuppressive therapy in this group (6,14).

The relationship between DGF and immunosuppression requires careful consideration. More intensive induction may reduce early rejection but can increase infection risk. In practice, this balance should be individualized according to patient risk.

Despite the exceptionally high prevalence of DGF in our cohort, long-term patient and graft outcomes remained favorable. Long-term follow-up revealed sustained patient survival, with estimated rates of 100% at 1 year, 97% at 3 years, 93% at 5 years, and 89% at 10 years. Death-censored graft survival remained comparably high, reaching 100%, 97%, 93%, and 86% at the respective time points. These outcomes indicate that favorable long-term patient and graft survival can be observed in recipients who develop DGF (2,6,7). The favorable survival observed in our cohort may reflect the combined effects of careful post-transplant monitoring, appropriate immunosuppressive management, and timely treatment of complications. Collectively, these findings support the concept that successful long-term outcomes can be achieved despite early graft dysfunction when recipients receive comprehensive post-transplant care. Prospective studies involving multiple transplant centers are necessary to validate the present findings and to clarify the recipient, -donor-, and transplant-related factors that may affect outcomes following DGF.

Study Limitations

The findings of this study should be interpreted in light of certain limitations. Its retrospective nature, conduct at a single transplant center, and modest cohort size may limit the applicability of the results to broader transplant populations. In addition, the absence of a comparator group without DGF precluded assessment of the independent effect of DGF on outcomes; therefore, our findings should be interpreted as descriptive rather than causal. Some potentially relevant variables, including donor instability and preservation details, were not consistently available. Therefore, our findings should be considered hypothesis-generating rather than confirmatory.

CONCLUSION

In conclusion, although DGF was highly prevalent in this deceased donor kidney transplant cohort, long-term patient and graft outcomes remained favorable. The occurrence of viral infectious complications highlights the importance of close monitoring and individualized immunosuppressive management. Therefore, careful post-transplant surveillance and tailored immunosuppressive strategies remain essential in recipients with DGF. These findings may help inform risk-adapted management approaches and support individualized care in this patient population.

Ethics

Ethics Committee Approval: The study protocol received approval from the Gazi University Ethics Commission before data extraction and analysis (approval no: 05.02.2026-E.1456672, date: 05.02.2026).

Informed Consent: All patient data were anonymized, and confidentiality was maintained throughout the study.

Footnotes

Authorship Contributions

Concept: S.K., H.S., R.K., A.D., Design: S.K., R.K., H.S., M.Y., A.D., Data Collection or Processing: H.S., R.K., R.K., A.D., Analysis or Interpretation: S.K., H.S., R.K., M.Y., Literature Search: H.S., R.K., S.K., M.Y., Writing: H.S., S.K., R.K., M.Y.

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

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

REFERENCES

1. Ponticelli C, Reggiani F, Moroni G. Delayed graft function in kidney transplant: risk factors, consequences and prevention strategies. J Pers Med. 2022; 12: 1557.
2. Yao Z, Kuang M, Li Z. Risk factors for delayed graft function in patients with kidney transplantation: a systematic review and meta-analysis. BMJ Open. 2025; 15: e087128.
3. Schröppel B, Legendre C. Delayed kidney graft function: from mechanism to translation. Kidney Int. 2014; 86: 251-8.
4. Nashan B, Abbud-Filho M, Citterio F. Prediction, prevention, and management of delayed graft function: where are we now? Clin Transplant. 2016; 30: 1198-208.
5. Tingle SJ, Figueiredo RS, Moir JA, Goodfellow M, Talbot D, Wilson CH. Machine perfusion preservation versus static cold storage for deceased donor kidney transplantation. Cochrane Database Syst Rev. 2019; 3: CD011671. Update in: Cochrane Database Syst Rev. 2024; 7: CD011671.
6. Li MT, Ramakrishnan A, Yu M, Daniel E, Sandra V, Sanichar N, et al. Effects of delayed graft function on transplant outcomes: a meta-analysis. Transplant Direct. 2023; 9: e1433.
7. Jayaram D, Kommareddi M, Sung RS, Luan FL. Delayed graft function requiring more than one-time dialysis treatment is associated with inferior clinical outcomes. Clin Transplant. 2012; 26: E536-43.
8. Helfer MS, Pompeo JC, Costa ORS, Vicari AR, Ribeiro AR, Manfro RC. Long-term effects of delayed graft function duration on function and survival of deceased donor kidney transplants. J Bras Nefrol. 2019; 41: 231-41.
9. Haberal M, Emiroglu R, Karakayali H, Torgay A, Moray G, Arslan G, et al. A corner-saving ureteral reimplantation technique without stenting. Transplant Proc. 2006; 38: 548-51.
10. Sözen H, Onaran M, Özen O, Dalgıç A. Urologic complications after renal transplant: a single-center experience. Exp Clin Transplant. 2017.
11. Mallon DH, Summers DM, Bradley JA, Pettigrew GJ. Defining delayed graft function after renal transplantation: simplest is best. Transplantation. 2013; 96: 885-9.
12. Roufosse C, Simmonds N, Clahsen-van Groningen M, Haas M, Henriksen KJ, Horsfield C, et al. A 2018 reference guide to the Banff classification of renal allograft pathology. Transplantation. 2018; 102: 1795-814. Erratum in: Transplantation. 2018; 102: e497. Erratum in: Transplantation. 2022; 106: e528.
13. Wu WK, Famure O, Li Y, Kim SJ. Delayed graft function and the risk of acute rejection in the modern era of kidney transplantation. Kidney Int. 2015; 88: 851-8.
14. Hirsch HH, Randhawa PS; AST Infectious Diseases Community of Practice. BK polyomavirus in solid organ transplantation-Guidelines from the American Society of Transplantation Infectious Diseases Community of Practice. Clin Transplant. 2019; 33: e13528.

Gazi Med J 2026;37(4):538-543

DOI: http://dx.doi.org/10.12996/gmj.2026.4765

Prognostic Significance of LODDS in Right-sided Colon Cancer: A Retrospective Analysis

Sağ Taraflı Kolon Kanserinde LODDS'nin Prognostik Önemi: Retrospektif Bir Analiz

Enes Aksu[1], Aydın Yavuz[2], Semih Cihan Yıldırım[2], Hasan Bostancı[1], Kürşat Dikmen[1], Hüseyin Göbüt[1], Ahmet Çağrı Büyükkasap[1], Sayın Altıner[1], Osman Yüksel[1]

[1]Department of General Surgery, Gazi University, Faculty of Medicine, Ankara, Türkiye
[2]Department of General Surgery, University of Health Sciences Türkiye, Sincan Training and Research Hospital, Ankara, Türkiye

ABSTRACT

Objective: The conventional tumor, node, metastasis (TNM) staging system for colorectal cancer possesses inherent limitations due to its dependence on lymph node dissection completeness. This study assessed the prognostic performance of the log odds of positive lymph nodes (LODDS) and lymph node ratio (LNR) against traditional TNM staging in patients who received curative resection for right-sided colon cancer.

Methods: We conducted a retrospective analysis of 132 patients who underwent curative resection for right-sided colon cancer between November 2014 and January 2020. LODDS and LNR were calculated and categorized using published cut-off values. Kaplan-Meier survival analysis and multivariable Cox proportional hazards regression identified independent predictors of 5-year mortality.

Results: During a median follow-up of 90.2 months, the 5-year overall survival rate was 50.2%. Multivariate Cox regression confirmed LODDS category, LNR category, AJCC stage, and patient age as independent predictors of 5-year mortality. Mortality hazard was markedly elevated for the highest-risk LODDS group [LODDS 2 versus LODDS 0: hazard ratio (HR) = 22.322; $p < 0.001$] compared to the highest LNR category (LNR 2 versus LNR 0: HR = 4.356; $p = 0.001$).

Conclusion: Both LODDS and LNR function as robust independent prognostic markers for 5-year mortality following curative resection for right-sided colon cancer. However, LODDS demonstrates superior discriminatory power in identifying patients at highest risk.

ÖZ

Amaç: Kolorektal kanserde geleneksel tümör, lenf düğümü, metastaz (TNM) evreleme sisteminin, lenf nodu diseksiyonunun yeterliliğine bağımlı olması nedeniyle bazı kısıtlılıkları bulunmaktadır. Bu çalışmada, sağ taraflı kolon kanseri nedeniyle küratif rezeksiyon uygulanan hastalarda Pozitif lenf nodlarının logaritmik odds oranı (LODDS) ve lenf nodu oranının (LNR) prognostik performansı geleneksel TNM evreleme sistemi ile karşılaştırılarak değerlendirildi.

Yöntemler: Kasım 2014 ile Ocak 2020 tarihleri arasında sağ taraflı kolon kanseri nedeniyle küratif rezeksiyon uygulanan 132 hastanın retrospektif analizi gerçekleştirildi. LODDS ve LNR değerleri hesaplandı ve daha önce yayımlanmış kesim noktaları kullanılarak kategorilere ayrıldı. Beş yıllık mortalitenin bağımsız belirleyicilerini saptamak amacıyla Kaplan-Meier sağkalım analizi ve çok değişkenli Cox orantısal riskler regresyon analizi uygulandı.

Bulgular: Ortanca 90,2 aylık takip süresinde beş yıllık genel sağkalım oranı %50,2 olarak bulundu. Çok değişkenli Cox regresyon analizi, LODDS kategorisi, LNR kategorisi, AJCC evresi ve hasta yaşının beş yıllık mortalitenin bağımsız belirleyicileri olduğunu doğruladı. En yüksek riskli LODDS grubunda mortalite riski, en yüksek LNR kategorisine kıyasla belirgin olarak daha yüksekti [LODDS 2 ve LODDS 0 karşılaştırması: risk oranı (HR) = 22,322; $p < 0,001$; LNR 2 ve LNR 0 karşılaştırması: HR = 4,356; $p = 0,001$].

Sonuç: Hem LODDS hem de LNR, sağ taraflı kolon kanserinde küratif rezeksiyon sonrasında beş yıllık mortalite açısından güçlü ve bağımsız

Cite this article as: Aksu E, Yavuz A, Yıldırım SC, Bostancı H, Dikmen K, Göbüt H, et al. Prognostic significance of LODDS in right-sided colon cancer: a retrospective analysis. Gazi Med J. 2026;37(4):538-543

Address for Correspondence/Yazışma Adresi: Enes Aksu, Department of General Surgery, Gazi University, Faculty of Medicine, Ankara, Türkiye
E-mail / E-posta: enesaksu@gazi.edu.tr
ORCID ID: orcid.org/0009-0002-8334-7618

Received/Geliş Tarihi: 10.07.2026
Accepted/Kabul Tarihi: 28.09.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

Incorporating LODDS into clinical assessment could enhance risk stratification and inform adjuvant treatment strategies.

Keywords: Right-sided colon cancer, log odds of positive lymph nodes (LODDS), lymph node ratio (LNR), prognosis, survival analysis

ÖZ

prognostik belirteçler olarak işlev görmektedir. Bununla birlikte, LODDS, en yüksek risk taşıyan hastaların belirlenmesinde daha üstün ayırt edici güce sahiptir. LODDS'nin klinik değerlendirmeye dahil edilmesi, risk sınıflandırmasının geliştirilmesine ve adjuvan tedavi stratejilerinin belirlenmesine katkı sağlayabilir.

Anahtar Sözcükler: Sağ taraflı kolon kanseri, pozitif lenf nodlarının logaritmik odds oranı (LODDS), lenf nodu oranı (LNR), prognoz, sağkalım analizi

INTRODUCTION

Colorectal cancer (CRC) remains one of the most prevalent malignancies and a major contributor to cancer-related deaths worldwide (1,2). Precise staging forms the cornerstone of prognostic evaluation and therapeutic planning. The American Joint Committee on Cancer (AJCC) tumor, node, metastasis (TNM) classification serves as the internationally recognized benchmark, with regional lymph node involvement (N-stage) consistently identified as the paramount determinant of survival (3).

Despite widespread adoption, the TNM N-stage classification has inherent limitations. Its principal constraint is dependency solely on the absolute number of positive lymph nodes. When inadequate lymphadenectomy yields fewer than the AJCC-recommended 12 nodes for pathological assessment, the risk of pathological under-staging increases, potentially leading to suboptimal treatment decisions (4). Furthermore, significant prognostic heterogeneity exists among patients within the same N-stage category when total nodes examined differ.

Alternative staging systems have been proposed to address these shortcomings. The lymph node ratio (LNR) represents the simple proportion of positive nodes relative to total retrieved nodes. The Log Odds of Positive Lymph Nodes (LODDS) offers a more sophisticated measure, defined by the formula: LODDS = log [(p + 0.5) / (n + 0.5)], where p is the number of positive nodes and n is the number of negative nodes. By integrating both components of nodal harvest, LODDS demonstrates superior statistical stability and prognostic reliability, especially in cases with limited lymph node yield (5,6).

The prognostic superiority of LODDS has been established across several malignancies, including gastric, breast, and rectal cancers (7-9). CRC is increasingly recognized not as a single entity, but as two distinct diseases: Right-sided and left-sided tumors, each with unique biological and clinical profiles. Right-sided colon cancers (RCC), originating from the cecum, ascending colon, and hepatic flexure, are frequently diagnosed at more advanced stages, exhibit distinct molecular characteristics, and are associated with poorer prognosis (10,11).

While the prognostic value of LODDS has been explored in pooled CRC cohorts and specifically in rectal cancer, its comparative performance against TNM staging and LNR has not been rigorously assessed in a cohort dedicated exclusively to RCC. The primary aim of this study was to evaluate the independent prognostic value of the LODDS system for long-term survival following curative resection for RCC, comparing its performance directly with conventional TNM stage and LNR.

MATERIALS AND METHODS

Study Design and Patient Cohort

We included patients who received an initial primary diagnosis of adenocarcinoma and subsequently underwent right hemicolectomy between November 2014 and January 2020. Right-sided colon tumors were anatomically defined as those located in the cecum, ascending colon, or hepatic flexure.

Inclusion Criteria:

(1) Primary pathological diagnosis of adenocarcinoma;

(2) History of undergoing right hemicolectomy;

(3) Achievement of curative-intent resection (R0);

(4) Minimum of 12 lymph nodes retrieved and examined postoperatively.

Exclusion Criteria:

(1) Familial adenomatous polyposis or Lynch syndrome;

(2) Age under 18 years;

(3) Synchronous tumors, peritoneal dissemination, or distant metastatic disease at diagnosis (Stage IV);

(4) Emergent surgery for perforation or obstruction;

(5) Death within the immediate postoperative period (first 30 days);

(6) Fewer than 12 lymph nodes harvested.

Parameters Investigated

We collected demographic information (gender, age), tumor-specific characteristics (location, size, differentiation grade, lymphovascular and perineural invasion), surgical details, and long-term survival outcomes.

Staging Systems Evaluated

Three distinct lymph node staging systems were compared:

1. TNM stage: Patients were staged according to AJCC 8th edition criteria (N0, N1a, N1b, N2a, N2b).

2. LNR: Calculated by dividing positive lymph nodes by total retrieved nodes. LNR values were stratified into three categories: LNR 0 (≤0.05), LNR 1 (>0.05 to ≤0.2), and LNR 2 (>0.2) (8).

3. Log odds of positive nodes (LODDS): Computed using the formula: log [(pnod + 0.5) / (tnod + 0.5)]. LODDS values were categorized into three groups: LODDS 0 (≤-1.36), LODDS 1 (>-1.36 to ≤-0.53), and LODDS 2 (>-0.53). These cut-off points were selected based on values demonstrated to hold prognostic significance in large-scale studies (5,8).

This single-center, retrospective cohort study was conducted at Gazi University, Faculty of Medicine Hospital. The study protocol adhered to the principles outlined in the Declaration of Helsinki and received approval from the Local Ethics Committee of Gazi University (approval number: E-77082166-604.01-1451101, date: 13/01/2026). Informed consent was obtained from all individual participants included in the study.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics 27. Continuous data are presented as mean ± standard deviation or median (range); categorical variables as frequencies (%). Normality was assessed with the Kolmogorov-Smirnov test. Group comparisons used the Independent Sample t-test (normal distribution) or the Mann-Whitney U test (non-normal). Pearson's χ^2 test analyzed qualitative variable associations. Survival was analyzed via Kaplan-Meier with Log-Rank test comparison. Independent prognostic factors were identified using Cox proportional hazards regression (Backward:LR method). The prognostic efficacy of TNM N-stage, LNR, and LODDS was compared using ROC curve analysis, calculating the AUC for each. A p-value < 0.05 was considered statistically significant.

RESULTS

The study cohort comprised 132 patients who underwent curative resection for right-sided colon cancer, consisting of 70 (53.0%) male and 62 (47.0%) female patients. Mean age at surgery was 64.11 ± 12.92 years. Most tumors were classified as moderately (55.3%) or well-differentiated (26.5%). According to AJCC staging, 43.9% were Stage 2 and 40.9% were Stage 3. The mean number of lymph nodes retrieved per patient was 35.36 ± 15.99, with a mean of 1.46 ± 3.09 metastatic lymph nodes identified. Median follow-up duration was 90.2 months, and the 5-year overall survival rate was 50.2%. The demographic and clinical characteristics of this patient population are detailed in Table 1.

Table 1. Baseline patient demographics and clinical characteristics.

Variable (n = 132)	n		%	
Gender				
Female	62		47.0	
Male	70		53.0	
Differentiation				
Poor	24		18.2	
Moderate	73		55.3	
Well	35		26.5	
Dukes category				
A	20		15.2	
B1	52		39.4	
B2	6		4.5	
C1	36		27.3	
C2	18		13.6	
Stage				
1	20		15.2	
2	58		43.9	
3	54		40.9	
LODDS				
0	86		65.2	
1	41		31.1	
2	5		3.7	
LNR				
0	103		78.0	
1	19		14.4	
2	9		6.8	
Quantitative variables		**Median**	**Min.**	**Max.**
Age at operation	64.11 ± 12.92	66.0	28.0	89.0
Number of resected lymph nodes	35.36 ± 15.99	31.0	12.0	91.0
Number of metastatic lymph nodes	1.46 ± 3.09	0.0	0.0	20.0
Tumor diameter (mm)	55.80 ± 25.36	50.0	2.0	130.0

LNR: Lymph node ratio, LODDS: Log odds of positive lymph nodes, Min.: Minimum, Max.: Maximum.

Univariate Analysis

Univariate analysis revealed no significant association between survival and gender or tumor differentiation (p > 0.05). However, survival was strongly correlated with Dukes Category (χ^2 = 25.955; p < 0.001) and AJCC Stage (χ^2 = 21.435; p < 0.001). Significant relationships were also found for LODDS Category (χ^2 = 12.426; p = 0.002), with LODDS 0 showing higher survival than categories 1-2, and for LNR Category (χ^2 = 12.525; p = 0.002). Deceased patients had significantly more metastatic lymph nodes (p < 0.001) and were older at surgery (p = 0.041) compared to survivors. No significant differences were observed in total retrieved lymph nodes or tumor size (p > 0.05). These results are summarized in Table 2.

Multivariate Analysis

In multivariate analysis, LODDS category, LNR category, AJCC Stage, and age at operation emerged as independent predictors of 5-year mortality (p < 0.05). After adjustment, LODDS category remained a significant prognostic factor. Compared to LODDS 0, the mortality hazard was 2.319 times higher for LODDS 1 [95% confidence interval (CI): 1.225-4.390; p = 0.010] and 22.322 times higher for LODDS 2 (95% CI: 6.552-76.052; p < 0.001). Similarly, LNR category was an independent prognostic factor, with LNR 1 [hazard ratio (HR) = 2.977; 95% CI: 1.463-6.060; p = 0.003] and LNR 2 (HR = 4.356; 95% CI: 1.774-10.693; p = 0.001) associated with increased risk versus LNR 0. AJCC stage also retained significance: Stage 2 (HR = 1.955; 95% CI: 1.235-3.095; p = 0.004) and Stage 3 (HR = 2.963; 95% CI: 1.952-4.497; p < 0.001) showed elevated mortality risk relative to Stage 1 (Table 3).

DISCUSSION

Accurate prognostic stratification is essential for adjuvant therapy decisions following resection for CRC. Our findings underscore a key limitation of conventional TNM N-stage: Its dependence on the absolute count of positive nodes, which renders it unreliable when lymph node retrieval is suboptimal (4). In contrast, this study

Table 2. Investigation of the relationship between general characteristics and survival status.

Survival variable	Survived (n = 85)		EX (n = 47)		Statistical analysis*
	n	%	n	%	
Gender					
Female	38	44.7	24	51.1	$\chi^2 = 0.491$
Male	47	55.3	23	48.9	p = 0.483
Differentiation					
Poor	15	17.6	9	19.1	$\chi^2 = 3.478$
Moderate	43	50.6	30	63.9	p = 0.176
Well	27	31.8	8	17.0	
Dukes category					
A	20	23.5	-	-	$\chi^2 = 25.955$
B1	38	44.7	14	29.8	**p < 0.001**
B2	3	3.5	3	6.4	
C1	19	22.4	17	36.2	
C2	5	5.9	13	27.6	
Stage					
1	20	23.6	-	-	$\chi^2 = 21.435$
2	41	48.2	17	36.2	**p < 0.001**
3	24	28.2	30	63.8	
LODDS					
0	64	75.3	22	46.8	$\chi^2 = 12.426$
1	20	23.5	21	44.7	**p = 0.002**
2	1	1.2	4	8.5	
LNR					
0	74	88.1	29	61.7	$\chi^2 = 12.525$
1	7	8.3	12	25.5	**p = 0.002**
2	3	3.6	6	12.8	

*Statistical analysis: Pearson-χ^2 (chi-square) cross tabulation was used to examine the relationship between two categorical variables. LNR: Lymph node ratio, LODDS: Log odds of positive lymph nodes.

demonstrates that both LNR and, more powerfully, LODDS provide robust independent prognostic value in right-sided colon cancer.

Multivariate Cox regression confirmed that LODDS category, LNR category, and AJCC stage are all independent predictors of 5-year mortality in RCC. This finding robustly supports our hypothesis that ratio-based staging systems, particularly LODDS and LNR, provide prognostic information comparable to—and potentially more robust than—traditional N-stage for right-sided colon cancers (5,6,12).

The principal finding of this study is the superior prognostic stratification offered by LODDS compared to LNR. The hazard for mortality in the highest LODDS category was 22.322 times greater than the lowest category, whereas the corresponding hazard for the highest LNR category was 4.356-fold. This enhanced performance is explained by the mathematical properties of the LODDS formula, which incorporates both positive and negative nodal counts into a continuous logarithmic metric, conferring greater stability against variations in total lymph node yield (5,6). Consequently, LODDS maintains prognostic accuracy in cohorts with heterogeneous lymphadenectomy, an advantage documented in colorectal and gastric cancers where it provides superior survival curve discrimination versus traditional N-staging (5,6,12). Its robustness as a prognostic indicator has been replicated in breast and other solid tumors (8,9,13,14). Our findings extend this evidence by specifically validating the high discriminatory capacity of LODDS in right-sided colon cancer.

Our analysis confirmed LNR as an independent prognostic factor, with significantly increased mortality hazard in its highest category. A critical limitation, however, is its inability to stratify the substantial subset of lymph node-negative patients, for whom LNR is uniformly zero. LODDS addresses this deficit by incorporating the denominator of negative nodes into a continuous variable, thereby unmasking prognostic heterogeneity within node-negative cohorts and offering more granular risk assessment than either N-stage or LNR (7,15,16). Although some reports suggest prognostic equivalence between LNR and LODDS (15,16), the broader literature, including larger series, indicates that LODDS provides superior discrimination, particularly for high-risk stratification (5,13,14).

Table 3. Independent risk factors associated with 5-year mortality (Cox regression analysis).

Variables	B	S.H.	Wald	SD	p	HR	95% confidence interval (HR)	
							Lower	Higher
LODDS 0*			26.511	2	**<0.001**			
LODDS 1	0.841	0.326	6.668	1	**0.010**	2.319	1.225	4.390
LODDS 2	3.106	0.625	24.654	1	**<0.001**	22.322	6.552	76.052
LNR 0*			15.535	2	**<0.001**			
LNR 1	1.091	0.363	9.054	1	**0.003**	2.977	1.463	6.060
LNR 2	1.471	0.458	10.311	1	**0,001**	4.356	1.774	10.693
Stage 1*			13.818	2	**<0.001**			
Stage 2	0.671	0.234	8.184	1	**0.004**	1.955	1.235	3.095
Stage 3	1.086	0.213	26.040	1	**<0.001**	2.963	1.952	4.497
Age at operation	0.065	0.015	18.122	1	**<0.001**	1.067	1.036	1.100

*: Reference category, SD: Standard deviation, HR: Hazard ratio, LNR: Lymph node ratio, LODDS: Log odds of positive lymph nodes.

Right- and left-sided colon cancers are now recognized as biologically and clinically distinct entities. Right-sided colon cancers are frequently diagnosed at more advanced stages, harbor distinct molecular profiles, and portend poorer prognosis (10,11). Much of the existing literature on LODDS has evaluated CRC as a homogeneous group or focused on rectal cohorts (14,15,17). A primary strength of this study is its exclusive focus on RCC, enabling precise assessment of LODDS and LNR within this specific anatomical subset. This approach is supported by work suggesting lymph node-related metrics are particularly influential for survival outcomes in right-sided tumors (11). Consequently, our findings indicate that LODDS can offer clinicians critical adjunctive data to personalize adjuvant therapy and identify high-risk patients within the RCC population.

Our finding that age at operation is an independent prognostic variable aligns with the established association between advanced age and inferior survival in RCC. This relationship may be mediated by higher prevalence of comorbidities, physiological frailty, and reduced tolerance to adjuvant regimens. Although the number of positive nodes retained its expected prognostic significance in univariate analysis, both LODDS and LNR demonstrated enhanced capacity for refined risk classification. Therefore, these results argue for clinical implementation of LODDS and LNR as complementary or alternative tools to N-stage, facilitating a more nuanced approach to adjuvant therapy planning and surveillance in RCC.

Study Limitations

This study has several limitations. Its retrospective, single-center design carries inherent biases and may limit generalizability. The statistical power for the highest-risk categories (LODDS 2 and LNR 2) is constrained by their small subgroup sizes. Although overall survival was the primary endpoint, disease-free survival data would have better assessed recurrence risk. Furthermore, our model did not incorporate established clinicopathological and molecular prognostic factors, such as microsatellite instability (MSI) status.

CONCLUSION

This study confirms that both LODDS and LNR are independent prognostic factors for 5-year mortality after curative resection of right-sided colon cancer, independent of AJCC Stage. Notably, LODDS demonstrated a superior ability to identify the highest-risk subgroup (LODDS 2), with a HR of 22.322. These results suggest that LODDS may serve as a valuable complement, or even a superior alternative, to the TNM system for prognostic assessment and personalization of adjuvant therapy. Incorporating LODDS into routine clinical practice could refine risk stratification and optimize treatment. Prospective multicenter studies are needed to validate these findings and integrate LODDS into standard decision-making algorithms and should incorporate disease-free survival as a primary endpoint alongside molecular markers such as MSI and KRAS mutation status, which were not assessed in the present study.

Ethics

Ethics Committee Approval: The study protocol adhered to the principles outlined in the Declaration of Helsinki and received approval from the Local Ethics Committee of Gazi University (approval number: E-77082166-604.01-1451101, date: 13/01/2026).

Informed Consent: Informed consent was obtained from all individual participants included in the study.

Footnotes

Preliminary results from this study were presented as a oral presentation at TSCRS, Antalya, Türkiye, May 16-20, 2025.

Authorship Contributions

Surgical and Medical Practice: K.D., O.Y., Concept: S.A., H.B., Design: E.A., O.Y., Data Collection or Processing: S.C.Y., K.D., Analysis or İnterpretation: E.A., A.Ç.B., Literature Search: S.C.Y., A.Y., Wrighting: E.A., S.A.

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

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

REFERENCES

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin. 2023; 73: 17-48.
2. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024; 74: 229-63.
3. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. CA Cancer J Clin. 2017; 67: 93-9.
4. Qiao Y, Zhu J, Han T, Jiang X, Wang K, Chen R, et al. Finding the minimum number of retrieved lymph nodes in node-negative colorectal cancer using real-world data and the SEER database. Int J Surg. 2023; 109: 4173-84.
5. Hu X, Jiang L, Wu J, Mao W. Prognostic value of log odds of positive lymph nodes, lymph node ratio, and N stage in patients with colorectal signet ring cell carcinoma: a retrospective cohort study. Front Surg. 2022; 9: 1019454.
6. Zhang W, Hou Z, Zhang L, Hong X, Wang W, Wu X, et al. A log odd of positive lymph nodes-based predictive model effectively forecasts prognosis and guides postoperative adjuvant chemotherapy duration in stage III colon cancer: a multi-center retrospective cohort study. BMC Cancer. 2024; 24: 1088.
7. Naidu K, Chapuis PH, Connell L, Chan C, Rickard M, Ng KS. Lymph node ratio prognosticates overall survival in patients with stage IV colorectal cancer. Tech Coloproctol. 2024; 28: 115.
8. Yavuz A, Buyukkasap C, Altiner S, Kurtoglu B, Koyuncuoglu C, Gobut H, et al. Log odds of positive nodes as a prognostic factor for rectal cancer: a retrospective study. Rev Assoc Med Bras (1992). 2024; 70: e20230417.
9. Zheng C, Huang J, Zhao S, Peng Y, Xiong Q, Li X, et al. Log odds of positive lymph nodes as a prognostic marker in resectable triple negative breast cancer. Sci Rep. 2025; 15: 30348.
10. Asghari-Jafarabadi M, Wilkins S, Plazzer JP, Yap R, McMurrick PJ. Prognostic factors and survival disparities in right-sided versus left-sided colon cancer. Sci Rep. 2024; 14: 12306.
11. Serafin M, Mąka M, Szostek J, Kania I, Jabłońska B, Mrowiec S. Side-specific prognostic factors in colon cancer: a retrospective analysis of right- and left-sided tumors. Cancers. 2025; 17: 3315.
12. Scarinci A, Di Cesare T, Cavaniglia D, Neri T, Colletti M, Cosenza G, et al. The impact of log odds of positive lymph nodes (LODDS) in colon

and rectal cancer patient stratification: a single-center analysis of 323 patients. Updates Surg. 2018; 70: 23-31.

13. Baqar AR, Wilkins S, Wang W, Oliva K, McMurrick P. Log odds of positive lymph nodes is prognostically equivalent to lymph node ratio in non-metastatic colon cancer. BMC Cancer. 2020; 20: 762.
14. Chen C, Xia HB, Yuan WW, Zhou MC, Zhang X, Xu AM. Developing a novel model for predicting overall survival in late-onset colon adenocarcinoma patients based on LODDS: a study based on the SEER database and external validation. Discov Oncol. 2025; 16: 99.
15. Qin ZF, Gao XY, Zhu YZ, Ma T, Chen GL, Ma DN, et al. Evaluation of LNR and modified N stage systems for prognostic stratification of metastatic lymph nodes in stage III colorectal Cancer. Sci Rep. 2025; 15: 13735.
16. Arrichiello G, Pirozzi M, Facchini BA, Facchini S, Paragliola F, Nacca V, et al. Beyond N staging in colorectal cancer: Current approaches and future perspectives. Front Oncol. 2022; 12: 937114.
17. Xu T, Zhang L, Yu L, Zhu Y, Fang H, Chen B, et al. Log odds of positive lymph nodes is an excellent prognostic factor for patients with rectal cancer after neoadjuvant chemoradiotherapy. Ann Transl Med. 2021; 9: 637.

Original Investigation | Özgün Araştırma

Gazi Med J 2026;37(4):544-550

DOI: http://dx.doi.org/10.12996/gmj.2026.4775

Persistent or Exacerbated Lymphopenia Predicts Mortality in Kidney Transplant Recipients Hospitalized with COVID-19

COVID-19 Geçiren Böbrek Nakli Alıcılarında Persistan veya Kötüleşen Lenfopeni Mortaliteyi Öngörür

Ali Can Kurtipek[1], Merve Öztürk[2], Ali Berkcan Bozdoğan[2], Ezgi Coşkun Yenigün[3], Emin Gemcioğlu[4], Osman İnan[2], Enes Seyda Şahiner[2], Oğuzhan Zengin[2], Selma Karaahmetoğlu[2], İhsan Ateş[2], Fatih Dede[3]

[1]Department of Internal Medicine, Ankara University, Faculty of Medicine, Ankara, Türkiye
[2]Clinic of Internal Medicine, University of Health Sciences Türkiye, Ankara Bilkent City Hospital, Ankara, Türkiye
[3]Clinic of Nephrology, University of Health Sciences Türkiye, Ankara Bilkent City Hospital, Ankara, Türkiye
[4]Clinic of Internal Medicine, University of Health Sciences Türkiye, Ankara Etlik City Hospital, Ankara, Türkiye

ABSTRACT

Objective: Kidney transplant recipients are at increased risk for severe coronavirus disease-2019 (COVID-19) because of chronic immunosuppressive therapy and impaired cellular immunity. Lymphopenia is a well-established predictor of adverse outcomes in the general population; however, the prognostic significance of persistent lymphopenia during hospitalization in kidney transplant recipients remains poorly defined. This study aimed to evaluate the association between persistent or exacerbated lymphopenia and clinical outcomes, including mortality and renal function, in hospitalized kidney transplant recipients with COVID-19.

Methods: We conducted a retrospective cohort study including adult kidney transplant recipients hospitalized with laboratory-confirmed COVID-19 at a tertiary referral center between March 2020 and January 2022. Patients were classified based on the presence or absence of persistent or exacerbated lymphopenia during hospitalization. Demographic characteristics, comorbidities, laboratory findings, treatments, and clinical outcomes were compared between groups. The primary outcome was in-hospital mortality. Secondary outcomes included intensive care unit (ICU) admission, requirement for respiratory support, requirement for invasive mechanical ventilation, length of hospital stay, requirement for hemodialysis, and renal function at 90 days after discharge. Independent predictors of mortality were identified using multivariable logistic regression analysis.

ÖZ

Amaç: Böbrek nakli alıcıları, kronik immünosüpresif tedavi ve bozulmuş hücresel immünite nedeniyle koronavirüs hastalığı-2019 (COVID-19) açısından ağır seyir riski taşıyan hasta grubunu oluşturmaktadır. Lenfopeninin genel popülasyonda kötü klinik sonuçların önemli bir belirteci olduğu bilinmesine karşın, hastanede yatış süresince devam eden veya kötüleşen lenfopeninin böbrek nakli alıcılarındaki prognostik değeri yeterince aydınlatılamamıştır. Bu çalışmada, persistan veya kötüleşen lenfopeni ile mortalite ve böbrek fonksiyonları başta olmak üzere klinik sonuçlar arasındaki ilişkinin değerlendirilmesi amaçlandı.

Yöntemler: Bu retrospektif kohort çalışmasına, Mart 2020 ile Ocak 2022 tarihleri arasında üçüncü basamak bir merkezde laboratuvar olarak doğrulanmış COVID-19 tanısıyla hastaneye yatırılan erişkin böbrek nakli alıcıları dahil edildi. Hastalar hastanede yatış sürecinde persistan veya kötüleşen lenfopeni varlığına göre gruplandırıldı. Demografik özellikler, komorbiditeler, laboratuvar bulguları, uygulanan tedaviler ve klinik sonuçlar gruplar arasında karşılaştırıldı. Birincil sonlanım noktası hastane içi mortaliteydi. İkincil sonlanım noktaları yoğun bakım ünitesi yatışı, solunum desteği gereksinimi, invaziv mekanik ventilasyon ihtiyacı, hastanede yatış süresi, hemodiyaliz gereksinimi ve taburculuk sonrası 90. gündeki böbrek fonksiyonları olarak belirlendi. Mortalite ile ilişkili bağımsız belirleyiciler çok değişkenli lojistik regresyon analizi ile değerlendirildi.

Cite this article as: Kurtipek AC, Öztürk M, Bozdoğan AB, Coşkun Yenigün E, Gemcioğlu E, İnan O, et al. Persistent or exacerbated lymphopenia predicts mortality in kidney transplant recipients hospitalized with COVID-19. Gazi Med J. 2026;37(4):544-550

Address for Correspondence/Yazışma Adresi: Ali Can Kurtipek, Department of Internal Medicine, Ankara University, Faculty of Medicine, Ankara, Türkiye
E-mail / E-posta: ackurtipek@ankara.edu.tr
ORCID ID: orcid.org/0000-0002-3504-7402

Received/Geliş Tarihi: 23.07.2026
Accepted/Kabul Tarihi: 12.08.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

Results: A total of 181 kidney transplant recipients were included, 63% of whom developed persistent or exacerbated lymphopenia. Patients with persistent or exacerbated lymphopenia experienced significantly higher rates of ICU admission, respiratory support, invasive mechanical ventilation, prolonged hospitalization, and in-hospital mortality compared with other patients (all $p < 0.05$). They also had significantly higher levels of inflammatory markers, including ferritin and C-reactive protein. In multivariable analysis, persistent/exacerbated lymphopenia remained independently associated with in-hospital mortality after adjustment for age, inflammatory markers, and hemodialysis requirement [adjusted odds ratio 14.16 (95% confidence interval: 3.6-55.7), $p < 0.001$]. However, persistent lymphopenia was not associated with loss of renal function at 90 days.

Conclusion: Persistent or exacerbated lymphopenia during hospitalization is a strong independent predictor of in-hospital mortality among kidney transplant recipients with COVID-19 and is associated with greater disease severity and increased resource utilization. In contrast, it does not appear to predict 90-day renal function decline. Serial assessment of lymphocyte counts may provide a simple, inexpensive, and readily available tool for early risk stratification in hospitalized kidney transplant recipients with COVID-19.

Keywords: COVID-19, kidney transplantation, lymphopenia

ÖZ

Bulgular: Çalışmaya toplam 181 böbrek nakli alıcısı dahil edildi ve bunların %63'ünde hastanede yatış süresince persistan veya kötüleşen lenfopeni saptandı. Persistan lenfopenisi bulunan hastalarda yoğun bakım ünitesine yatış, solunum desteği gereksinimi, invaziv mekanik ventilasyon ihtiyacı, hastanede kalış süresi ve hastane içi mortalite oranları anlamlı olarak daha yüksekti (tüm karşılaştırmalar için $p < 0,01$). Ayrıca ferritin ve C-reaktif protein düzeyleri de bu grupta anlamlı olarak daha yüksekti. Çok değişkenli analizde persistan veya kötüleşen lenfopeni; yaş, enflamatuvar belirteçler ve hemodiyaliz gereksinimi için düzeltme yapıldıktan sonra da hastane içi mortalitenin bağımsız bir belirleyicisi olarak saptandı [düzeltilmiş olasılık oranı: 14,16 (%95 güven aralığı: 3,6-55,7), $p < 0,001$]. Buna karşın, persistan lenfopeni ile 90. gündeki böbrek fonksiyon kaybı arasında anlamlı ilişki saptanmadı.

Sonuç: Hastanede yatış süresince devam eden veya kötüleşen lenfopeni, COVID-19 nedeniyle hastaneye yatırılan böbrek nakli alıcılarında hastane içi mortalitenin güçlü ve bağımsız bir belirleyicisidir ve daha ağır hastalık seyri ile artmış sağlık hizmeti gereksinimiyle ilişkilidir. Buna karşın, orta dönem böbrek fonksiyon kaybını öngörmemektedir. Seri mutlak lenfosit sayısı değerlendirmesi, hastaneye yatırılan böbrek nakli alıcılarında erken risk sınıflandırmasında geleneksel enflamatuvar ve renal belirteçleri tamamlayıcı, kolay uygulanabilir ve düşük maliyetli bir yöntem olarak kullanılabilir.

Anahtar Sözcükler: COVID-19, böbrek nakli, lenfopeni

INTRODUCTION

Coronavirus disease-2019 (COVID-19), caused by severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2), has caused substantial global morbidity and mortality since its emergence in late 2019. Although widespread vaccination, antiviral therapies, and advances in supportive care have improved clinical outcomes over the course of the pandemic, immunocompromised individuals remain at increased risk for severe disease, hospitalization, and death. Among these populations, kidney transplant recipients (KTRs) represent a particularly vulnerable group because of chronic immunosuppressive therapy, impaired cellular immunity, and the high prevalence of comorbid conditions such as hypertension, diabetes mellitus, and chronic kidney disease (CKD) (1,2).

COVID-19 frequently affects the kidneys through multiple mechanisms, including direct viral cytotoxicity, dysregulated immune activation, endothelial injury, thrombotic microangiopathy, and hemodynamic alterations. Consequently, acute kidney injury (AKI) is a common complication of COVID-19 and has consistently been associated with increased mortality and adverse long-term renal outcomes (3-6). In KTRs, the risk of graft dysfunction may be further exacerbated by baseline allograft vulnerability, nephrotoxic medications, and modifications to immunosuppressive therapy during acute infection (2,7). Although many patients recover from the acute phase of COVID-19, persistent deterioration of renal function has been reported after hospital discharge, which emphasizes the importance of identifying prognostic factors associated with subsequent graft outcomes (5).

Lymphopenia is one of the most consistent hematological abnormalities observed in patients with COVID-19 and is considered a hallmark of dysregulated host immune responses. Several studies have demonstrated that reduced peripheral lymphocyte counts are associated with greater disease severity, respiratory failure, intensive care unit (ICU) admission, and mortality in the general population (1,8). The proposed mechanisms include direct virus-induced lymphocyte apoptosis, cytokine-mediated immune exhaustion, bone marrow suppression, and redistribution of activated lymphocytes to infected tissues (8). In KTRs, interpretation of lymphocyte counts is inherently more complex because chronic immunosuppressive therapy may induce baseline lymphopenia, impair adaptive immune responses, and delay immune reconstitution following viral infection (2).

Although numerous studies have described the clinical characteristics and outcomes of COVID-19 in KTRs, relatively few have specifically evaluated the prognostic significance of persistent lymphopenia or the trajectory of lymphocyte recovery during hospitalization in relation to both patient survival and renal allograft outcomes (2,7-10). Most previous studies have focused on lymphocyte counts at hospital admission, whereas the clinical significance of persistent or worsening lymphopenia throughout hospitalization remains poorly defined. We therefore hypothesized that failure of lymphocyte recovery reflects sustained immune dysfunction and may provide prognostic information beyond a single admission measurement.

Accordingly, the aim of this study was to evaluate whether persistent or worsening lymphopenia during hospitalization is associated with adverse clinical outcomes, in-hospital mortality, and 90-day renal function among KTRs hospitalized with COVID-19.

MATERIALS AND METHODS

Study Design and Population

This retrospective, observational cohort study included adult KTRs hospitalized with laboratory-confirmed COVID-19 at University

of Health Sciences Türkiye, Ankara Bilkent City Hospital Hospital between March 2020 and January 2022. A COVID-19 diagnosis was confirmed by positive reverse transcriptase-polymerase chain reaction testing for severe acute respiratory SARS-CoV-2.

A total of 211 consecutive COVID-19 KTRs were screened. Patients were excluded if they had active malignancy, pregnancy, severe chronic lung disease, congenital heart disease, and end-stage kidney disease requiring maintenance dialysis prior to admission. After applying the exclusion criteria, 181 patients were included in the final analysis.

The study protocol was approved by the University of Health Sciences Türkiye, Ankara City Hospital Clinical Research Ethics Committee no. 1 (approval no: E1-22-2292, date: 12.01.2022). The requirement for informed consent was waived because of the retrospective nature of the study.

Definitions

Admission lymphopenia was defined as either (1) an absolute lymphocyte count (ALC) <1,000 cells/µL at hospital admission or (2) an ALC accompanied by a decline of more than 50% compared with the patient's pre-infection baseline lymphocyte count. Transient lymphopenia was defined as recovery of the ALC to >1,000 cells/µL within 10 days after COVID-19 diagnosis.

Persistent lymphopenia was defined as failure of the ALC to recover above 1,000 cells/µL beyond 10 days after diagnosis or until death, if death occurred between 7 and 10 days.

Exacerbated lymphopenia was defined as the development of lymphopenia during hospitalization in patients who did not meet the admission lymphopenia criteria, with lymphopenia persisting for at least 7 consecutive days.

Patients who died within the first 7 days of admission were excluded from the lymphopenia groups. Baseline renal function was determined using the most recent outpatient serum creatinine measurement obtained before COVID-19 diagnosis. Renal funcion loss at 90 days were calculated using nearest serum creatinine measurement to 90th day (up to 30 days before and after the 90th day), if available. Estimated glomerular filtration rates (eGFR) were calculated using the CKD-Epidemiology Collaboration creatinine equation. A 25% or greater loss in eGFR was considered a loss of at 90 days.

Data Collection

Electronic medical records were reviewed to obtain demographic characteristics, transplant-related variables, comorbidities, maintenance immunosuppressive therapy, clinical presentation, laboratory findings at admission and during hospitalization, COVID-19-specific treatments, oxygen supplementation requirements, ICU admission, mechanical ventilation, duration of hospitalization, and in-hospital mortality.

Renal outcomes included the development of AKI, the requirement for renal replacement therapy, serum creatinine levels, and eGFR, assessed during hospitalization and follow-up. Changes in immunosuppressive treatment during hospitalization were also recorded.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics version 27.0 (IBM Corp., Armonk, NY, USA). Continuous variables were assessed for normality using histogram graphs and Kolmogorov-Smirnov test and presented as mean ± standard deviation. Categorical variables are presented as frequencies and percentages. Comparisons between groups were performed using the Independent Samples t-test for normally distributed continuous variables, the Mann-Whitney U test for non-normally distributed continuous variables, and the chi-square test or Fisher's exact test for categorical variables, as appropriate.

Factors associated with in-hospital mortality were evaluated using binary logistic regression. Variables considered clinically relevant and demonstrating an association with mortality in univariate analysis ($p < 0.10$) were selected as candidates for multivariable analysis. To minimize overfitting, the number of variables in the multivariable model was restricted according to the events-per-variable principle. Age was included in the multivariable model regardless of its significance in univariate analysis because of its established clinical relevance. Variables were checked for multicollinearity using variance inflation factors (VIF); VIF <2 was considered acceptable. Variable selection for the final model was performed using backward-stepwise likelihood-ratio elimination; adjusted odds ratios (ORs) with 95% confidence intervals (CIs) were reported.

All statistical tests were two-sided, and a p-value <0.05 was considered statistically significant.

RESULTS

A total of 181 KTRs hospitalized with COVID-19 were included in the analysis. Five patients died before 7th day of their hospitalization, and were excluded from lymphopenia groups. Persistent or exacerbated lymphopenia was observed in 111 (63.1%) patients, whereas 65 (36.9%) patients had transient or no lymphopenia. Patients with persistent/exacerbated lymphopenia were significantly older than those without persistent lymphopenia (53.1 ± 12.1 vs. 48.6 ± 12.4 years, $p = 0.018$) and had a higher prevalence of hypertension (76.6% vs. 52.3%, $p < 0.001$). All patients were on maintenance immunosuppressive therapy. Baseline immunosuppressive regimens, years since transplantation, sex distribution, and the prevalence of diabetes mellitus and coronary artery disease were comparable between groups. Symptoms of COVID-19 were comparable between the lymphopenia groups. Corticosteroid treatment for COVID-19 was more prevalent among the persistent or exacerbated lymphopenia group (Table 1).

Baseline renal function was comparable between groups. Patients with persistent/exacerbated lymphopenia had significantly higher serum ferritin levels (866.1 ± 901.7 vs. 532.0 ± 633.3 ng/mL, $p = 0.009$), lower ALCs at admission (634.8 ± 424.1 vs. 920.5 ± 513.4 cells/µL, $p < 0.001$), and lower monocyte and eosinophil counts. No significant differences were observed in serum creatinine, estimated glomerular filtration rate (eGFR), C-reactive protein (CRP), or baseline lymphocyte counts (Table 2).

Persistent or exacerbated lymphopenia was associated with significantly worse clinical outcomes. Compared with patients with transient or no lymphopenia, these patients had higher rates of ICU

admission (51.4% vs. 18.5%, p < 0.001), invasive or non-invasive mechanical ventilation (29.7% vs. 9.2%, p = 0.002), intubation (22.5% vs. 6.2%, p = 0.005), and any respiratory support (84.7% vs. 61.5%, p < 0.001). Hospital length of stay was also significantly longer (19.1 ± 18.9 vs. 12.2 ± 10.2 days, p < 0.001). In-hospital mortality was markedly higher among patients with persistent or exacerbated lymphopenia (35.1% vs. 7.7%; p < 0.001). However, 90-day renal function loss was comparable between the groups (10.4% vs. 10.9%, p = 1.000) (Table 3).

In univariate logistic regression analysis, hemodialysis requirement, persistent or exacerbated lymphopenia, corticosteroid therapy for COVID-19, higher admission levels of CRP, ferritin, and urea, increased neutrophil count, and older age were significantly associated with in-hospital mortality. In the multivariable model, hemodialysis requirement (OR: 29.38, 95% CI: 6.61-130.64, p < 0.001); persistent/exacerbated lymphopenia (OR: 14.16, 95% CI: 3.60-55.69, p < 0.001); admission CRP (OR: 1.009 per mg/L increase, 95% CI: 1.003-1.015, p = 0.005); and age (OR: 1.044 per year, 95% CI: 1.00-1.083, p = 0.023) remained independent predictors of in-hospital mortality (Table 4).

Table 1. Comparison of demographics, comorbidities, baseline treatments, symptoms, and COVID-19 therapies between the groups.

	Persistent or exacerbated lymphopenia (n = 111)	Transient or no lymphopenia (n = 65)	p
Age (years)	53.1 ± 12.1	48.6 ± 12.4	0.018†
Male sex	78 (70.3%)	49 (75.4%)	0.465*
Diabetes mellitus	37 (33.3%)	20 (30.8%)	0.726*
Hypertension	85 (76.6%)	34 (52.3%)	<0.001*
Coronary artery disease	24 (21.6%)	13 (20.0%)	0.799*
Years since transplantation	7.2 ± 6.07	6.32 ± 4.97	0.42‡
Baseline immunosuppressive therapies			
Corticosteroids	103 (92.8%)	63 (96.9%)	0.327#
Tacrolimus	83 (74.8%)	56 (86.2%)	0.074*
Micofenilat mofetil	81 (73.0%)	54 (83.1%)	0.126*
Everolimus	11 (9.9%)	6 (9.2%)	0.883*
Cyclosporin	11 (9.9%)	3 (4.6%)	0.21*
Azathiopurine	8 (7.2%)	2 (3.1%)	0.327#
Sirolimus	4 (3.6%)	3 (4.6%)	0.71#
Admission symptoms			
Non-specific	59 (53.2%)	32 (49.2%)	0.615*
Fever	71 (64.0%)	32 (49.2%)	0.056*
Respiratory	79 (71.2%)	38 (58.5%)	0.085*
Gastrointestinal	23 (20.7%)	17 (26.2%)	0.406*
Treatments for COVID-19			
Corticosteroids	91 (82.0%)	35 (53.8%)	<0.001*
Colchicine	43 (38.7%)	19 (29.2%)	0.203*
Plasmapheresis	3 (2.7%)	0 (0.0%)	0.297#
IV-immunoglobulin	3 (2.7%)	0 (0.0%)	0.3#

†: Independent Samples t-test, *: Chi-squared test, ‡: Mann-Whitney U test, #: Fisher's exact test, COVID-19: Coronavirus disease-2019, IV: Intravenous.

Table 2. Comparison of laboratory variables between the groups.

	Persistent or exacerbated lymphopenia (n = 111)	Transient or no lymphopenia (n = 65)	p
Baseline eGFR (mL/dk/1.73 m²)	58.7 ± 22.2	60.8 ± 18.6	0.558†
Admission eGFR (mL/dk/1.73 m²)	49.3 ± 27.1	49.7 ± 20.2	0.9†
Urea (mg/dL)	79.9 ± 59.3	72.2 ± 50	0.871‡
Creatinine (mg/dL)	1.99 ± 1.34	1.77 ± 0.89	0.966‡
Albumin (g/L)	36.7 ± 6.29	38.3 ± 6.1	0.105†
Ferritin (mg/dL)	866.1 ± 901.7	532 ± 633.3	0.009‡
C-reactive protein (mg/dL)	84.6 ± 68.7	74.8 ± 76	0.12‡
White blood cells (/µL)	7195.7 ± 5111.5	7893.4 ± 4895.8	0.159‡
Neutrophils (/µL)	5916.1 ± 4946.3	6090.9 ± 4625.9	0.569‡
Monocytes (/µL)	411.4 ± 323	580 ± 987.9	0.032‡
Eosinophils (/µL)	29.5 ± 42.1	54.1 ± 76.2	0.033‡
Hemoglobin (g/dL)	11.8 ± 2.7	12.5 ± 2.07	0.061†
Platelets (x10⁹/L)	199.6 ± 79.6	223.3 ± 101.9	0.089†
Lymphocytes (/µL)	634.8 ± 424.1	920.5 ± 513.4	<0.001‡
Baseline lymphocytes (/µL)	1600.8 ± 817.3	1735.6 ± 821.1	0.212‡

†: Independent Samples t-test, ‡: Mann-Whitney U test, eGFR: Estimated glomerular filtration rate.

Table 3. Comparison of clinical outcomes between the groups.

	Persistent or exacerbated lymphopenia (n = 111)	Transient or no lymphopenia (n = 65)	p
Mortality	39 (35.1%)	5 (7.7%)	<0.001*
Mortality time (days)	21.5 ± 17.6	24.4 ± 14.5	0.591‡
Length-of-stay	19.1 ± 18.9	12.2 ± 10.2	<0.001‡
Length of ICU stay	17.1 ± 21.6	15 ± 11.8	0.975‡
ICU admission	57 (51.4%)	12 (18.5%)	<0.001*
Hemodialysis	13 (11.7%)	7 (10.8%)	0.849*
Renal function loss (at 90 days)	7 (10.4%)	5 (10.9%)	1#
Intubation	25 (22.5%)	4 (6.2%)	0.005*
Mechanical ventilation	33 (29.7%)	6 (9.2%)	0.002*
Any respiratory support	94 (84.7%)	40 (61.5%)	<0.001*

*: Chi-squared test, ‡: Mann-Whitney U test, #: Fisher's exact test, ICU: Intensive care unit.

Table 4. Logistic regression models predicting mortality.

	Univariate logistic regression		Multivariate logistic regression	
	OR (95% CI)	**p**	**OR (95% CI)**	**p**
Hemodialysis	13.49 (4.629-39.33)	<0.001	29.38 (6.608-130.64)	<0.001
P/E lymphopenia	7.333 (2.735-19.67)	<0.001	14.16 (3.598-55.69)	<0.001
Corticosteroid therapy for COVID-19	2.848 (1.68-4.829)	<0.001		
CRP at admission	1.009 (1.005-1.014)	<0.001	1.009 (1.003-1.015)	0.005
Ferritin at admission	1.001 (1-1.001)	<0.001		
Neutrophils at admission	1 (1-1)	<0.001		
Urea at admission	1.001 (1-1.002)	<0.001		
Age	1.048 (1.018-1.079)	0.002	1.044 (1.006-1.083)	0.023

P/E: Persistent or exacerbated, CRP: C-reactive protein, OR: Odds ratio, CI: Confidence interval, COVID-19: Coronavirus disease-2019.

DISCUSSION

In this retrospective cohort of KTRs hospitalized with COVID-19, we demonstrated that persistent or exacerbated lymphopenia was strongly associated with adverse clinical outcomes, including ICU admission, need for respiratory support, invasive mechanical ventilation, prolonged hospitalization, and in-hospital mortality. More importantly, persistent or exacerbated lymphopenia remained an independent predictor of mortality after adjustment for age, inflammatory markers, and requirement for hemodialysis. In contrast, persistent lymphopenia was not associated with loss of renal function 90 days after infection. Collectively, these findings suggest that failure of lymphocyte recovery primarily reflects sustained immune dysfunction and disease severity, rather than subsequent deterioration of kidney allograft function.

KTRs remain among the most vulnerable populations to severe COVID-19 because of chronic immunosuppressive therapy, impaired T-cell-mediated immunity, and a high prevalence of cardiovascular and metabolic comorbidities. Early reports from international transplant registries consistently demonstrated markedly higher rates of hospitalization, ICU admission, and mortality among KTRs than those in the general population (11-14). Unlike previous studies that primarily evaluated lymphocyte counts at hospital admission, our study specifically focused on the longitudinal trajectory of lymphocyte recovery during hospitalization. This dynamic approach may better capture persistent immune dysfunction and may explain why serial lymphocyte assessment provided prognostic information beyond a single baseline measurement. Although widespread vaccination, antiviral therapies, and improvements in supportive care have substantially altered the clinical course of COVID-19, KTRs continue to experience disproportionately poor outcomes because vaccine-induced immune responses are frequently attenuated and protection against emerging viral variants remains incomplete. Recent evidence further suggests that mortality among KTRs has remained considerably higher than that of the general population even during the Delta and Omicron eras, underscoring the continued importance of accurate risk stratification in this vulnerable population (15).

Lymphopenia is hallmark laboratory abnormality observed in COVID-19 and has consistently been associated with severe disease and death in the general population (1,16). Proposed mechanisms include direct viral injury to lymphocytes, cytokine-mediated apoptosis, immune exhaustion, impaired lymphopoiesis, and redistribution of activated lymphocytes into infected tissues (16,17). Among KTRs, these mechanisms are likely to be amplified by chronic immunosuppressive therapy, particularly antimetabolites and corticosteroids, leading to impaired lymphocyte proliferation and delayed adaptive immune recovery. Consequently, persistent lymphopenia may identify patients in whom immune recovery remains inadequate despite resolution of the initial viral insult.

Our findings extend previous observations by demonstrating that not only lymphopenia at hospital admission but also the persistence or worsening of lymphopenia throughout hospitalization identify patients at particularly high risk for death. While admission lymphocyte counts provide a static assessment of immune status, serial lymphocyte monitoring more accurately reflects the dynamic interaction between viral clearance, host immune recovery, systemic inflammation, and ongoing immunosuppression. Consistent with this concept, persistent or exacerbated lymphopenia remained independently associated with mortality after adjustment for age, CRP, and the requirement for hemodialysis. Moreover, recent evidence has shown that even early lymphopenia independently predicts adverse outcomes among hospitalized patients with COVID-19, supporting the hypothesis that persistent lymphocyte depletion represents an even stronger indicator of sustained immune dysfunction (18).

Our findings also suggest that persistent lymphopenia may represent an integrated biomarker reflecting both disease severity and inadequate recovery of adaptive immunity in KTRs. During severe COVID-19, antimetabolite therapy is frequently reduced or withdrawn temporarily to facilitate immune recovery, whereas calcineurin inhibitors and corticosteroids are generally individualized based on clinical status and graft function. Nevertheless, recovery of lymphocyte counts remains highly variable among transplant recipients, suggesting that persistent lymphopenia reflects ongoing immune dysregulation rather than immunosuppressive therapy alone. Contemporary reviews have similarly emphasized that balancing the restoration of antiviral immunity with minimizing the risk of acute rejection remains one of the principal therapeutic challenges in KTRs with COVID-19 (19).

Consistent with previous studies, markers of systemic inflammation were also associated with adverse outcomes in our cohort. Patients with persistent or exacerbated lymphopenia had significantly higher ferritin concentrations, whereas elevated CRP on admission independently predicted mortality. Hyperinflammation and immune dysregulation are well-recognized features of severe COVID-19 and contribute to endothelial injury, microvascular thrombosis, and multiorgan dysfunction (1,16,20). The coexistence of elevated inflammatory markers and persistent lymphopenia likely reflects a maladaptive immune response characterized by excessive innate immune activation together with impaired adaptive immunity, consistent with the cytokine storm paradigm proposed early during the pandemic (20).

The requirement for hemodialysis emerged as the strongest independent predictor of mortality in our cohort. AKI is a common complication of severe COVID-19 and has consistently been associated with increased mortality (2-6). The underlying mechanisms are multifactorial and include systemic inflammation, ischemic tubular injury, endothelial dysfunction, complement activation, and direct or indirect viral effects on the kidney (3,5,21). In KTRs, these mechanisms may be further aggravated by pre-existing allograft dysfunction, calcineurin inhibitor nephrotoxicity, volume depletion, and modifications of immunosuppressive therapy during acute illness. Our findings therefore reinforce previous observations that severe renal impairment requiring kidney replacement therapy is a robust marker of critical illness and poor prognosis among KTRs.

Study Limitations

This study has several limitations. First, its retrospective single-center design may have introduced selection bias and limited the generalizability of our findings to other transplant populations and healthcare settings. Second, although multivariable analyses were performed to adjust for clinically relevant covariates, the observational nature of the study precludes causal inference, and residual confounding from unmeasured variables cannot be excluded. Third, the limited number of mortality events may have reduced the statistical power of multivariable models and increased the risk of model overfitting, despite adherence to recommended model-building strategies. Furthermore, lymphocyte subset analysis was unavailable, preventing the evaluation of whether specific immune cell populations contributed differentially to the observed associations. Finally, the relatively small number of renal outcome events may have limited our ability to detect modest associations between persistent lymphopenia and subsequent allograft function.

CONCLUSION

Persistent or worsening lymphopenia during hospitalization strongly and independently predicts in-hospital mortality among KTRs with COVID-19 and is associated with greater disease severity, increased respiratory support requirements, and prolonged hospitalization. In contrast, persistent lymphopenia in our cohort was not associated with loss of renal function at 90 days, suggesting that sustained lymphocyte depletion primarily reflects acute immune dysregulation rather than subsequent allograft deterioration. Because ALCs are inexpensive, routinely available, and easily and serially monitored, they may serve as practical adjuncts to conventional inflammatory and renal biomarkers for early risk stratification in hospitalized KTRs. Prospective studies with longer follow-up are warranted to validate these findings and to determine whether therapeutic strategies aimed at improving immune recovery can translate into better patient and allograft outcomes.

Ethics

Ethics Committee Approval: The study protocol was approved by the University of Health Sciences Türkiye, Ankara Bilkent City Hospital Clinical Research Ethics Committee no. 1 (approval no: E1-22-2292, date: 12.01.2022).

Informed Consent: The requirement for informed consent was waived because of the retrospective nature of the study.

Acknowledgments

During the preparation of this article, the authors utilized artificial intelligence for language editing. The authors bear full responsibility for the accuracy, completeness, and scientific content of the final version of the article.

Footnotes

Authorship Contributions

Concept: A.C.K., M.Ö., A.B.B., E.C.Y., E.G., O.İ., E.S.Ş., O.Z., S.K., İ.A., F.D., Design: A.C.K., M.Ö., Data Collection or Processing: A.C.K., M.Ö., A.B.B., Analysis or Interpretation: A.C.K., E.C.Y., E.G., O.İ., E.S.Ş., O.Z., S.K., İ.A., F.D., Literature Search: A.C.K., M.Ö., Writing: A.C.K., M.Ö., A.B.B., E.C.Y., E.G., O.İ., E.S.Ş., O.Z., S.K., İ.A., F.D.

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

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

REFERENCES

1. Gao YD, Ding M, Dong X, Zhang JJ, Kursat Azkur A, Azkur D, et al. Risk factors for severe and critically ill COVID-19 patients: a review. Allergy. 2021; 76: 428-55.
2. Cravedi P, Mothi SS, Azzi Y, Mella A, Farouk S, Doucette JT, et al. COVID-19 and kidney transplantation: results from the TANGO international transplant consortium. Am J Transplant. 2020; 20: 3140-8.
3. Chan L, Chaudhary K, Saha A, Chauhan K, Vaid A, Baweja M, et al. Acute kidney injury in hospitalized patients with COVID-19. medRxiv [Preprint]. 2020: 8: 2020.05.04.20090944. Update in: J Am Soc Nephrol. 2021; 32: 151-60.
4. Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. Kidney Int. 2020; 97: 829-38.
5. Nugent J, Aklilu A, Yamamoto Y, Simonov M, Li F, Biswas A, Ghazi L, et al. Assessment of acute kidney injury and longitudinal kidney function after hospital discharge among patients with and without COVID-19. JAMA Netw Open. 2021; 4: e211095.
6. Akilesh S, Nast CC, Yamashita M, Henriksen K, Charu V, Troxell ML, et al. Multicenter clinicopathologic correlation of kidney biopsies performed in COVID-19 patients presenting with acute kidney injury or proteinuria. Am J Kidney Dis. 2021; 77: 82-93.e1.
7. Caillard S, Anglicheau D, Matignon M, Durrbach A, Greze C, Frimat L, et al.; French SOT COVID Registry. An initial report from the French SOT COVID registry suggests high mortality due to COVID-19 in recipients of kidney transplants. Kidney Int. 2020; 98: 1549-58.

8. Tan L, Wang Q, Zhang D, Ding J, Huang Q, Tang YQ, et al. Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. Signal Transduct Target Ther. 2020; 5: 33. Erratum in: Signal Transduct Target Ther. 2020; 5: 61.
9. Monfared A, Akhondzadeh L, Mousazadeh M, Jafari A, Khosravi M, Lebadi M, et al. COVID-19 in renal transplant recipients and general population: a comparative study of clinical, laboratory, and radiological features, severity, and outcome. Virol J. 2021; 18: 243.
10. Karruli A, Spiezia S, Boccia F, Gagliardi M, Patauner F, Salemme A, et al. Effect of immunosuppression maintenance in solid organ transplant recipients with COVID-19: systematic review and meta-analysis. Transpl Infect Dis. 2021; 23: e13595.
11. Akalin E, Azzi Y, Bartash R, Seethamraju H, Parides M, Hemmige V, et al. Covid-19 and kidney transplantation. N Engl J Med. 2020; 382: 2475-7.
12. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature. 2020; 584: 430-6.
13. Kates OS, Haydel BM, Florman SS, Rana MM, Chaudhry ZS, Ramesh MS, et al.; UW COVID-19 SOT Study Team. Coronavirus disease 2019 in solid organ transplant: a multicenter cohort study. Clin Infect Dis. 2021; 73: e4090-9.
14. Pereira MR, Mohan S, Cohen DJ, Husain SA, Dube GK, Ratner LE, et al. COVID-19 in solid organ transplant recipients: initial report from the US epicenter. Am J Transplant. 2020; 20: 1800-8.
15. Smith B, Nair S, Wadei H, Mai M, Khamash H, Schinstock C, et al. Increased mortality in kidney transplant recipients during the delta/omicron era of the COVID-19 pandemic despite widespread vaccination. Clin Transplant. 2025; 39: e70071.
16. Huang I, Pranata R. Lymphopenia in severe coronavirus disease 2019 (COVID-19): systematic review and meta-analysis. J Intensive Care. 2020; 8: 36.
17. Diao B, Wang C, Tan Y, Chen X, Liu Y, Ning L, et al. Reduction and functional exhaustion of t cells in patients with coronavirus disease 2019 (COVID-19). Front Immunol. 2020; 11: 827.
18. Okada K, Tagami T, Otaguro T, Hayakawa M, Yamakawa K, Endo A, et al. Early lymphopenia as a predictor of COVID-19 outcomes: a multicenter cohort study. Acute Med Surg. 2025; 12: e70044.
19. Azzi Y, Bartash R, Scalea J, Loarte-Campos P, Akalin E. COVID-19 and solid organ transplantation: a review article. Transplantation. 2021; 105: 37-55.
20. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ; HLH Across Speciality Collaboration, UK. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet. 2020; 395: 1033-4.
21. Naicker S, Yang CW, Hwang SJ, Liu BC, Chen JH, Jha V. The novel coronavirus 2019 epidemic and kidneys. Kidney Int. 2020; 97: 824-8.

Original Investigation | Özgün Araştırma

Gazi Med J 2026;37(4):551-554

DOI: http://dx.doi.org/10.12996/gmj.2026.4779

Seasonal Variation in Histologic and Endoscopic Activity of Pediatric Eosinophilic Esophagitis: A Repeated Endoscopy Cohort

Pediatrik Eozinofilik Özofajitte Histolojik ve Endoskopik Aktivitenin Mevsimsel Değişimi: Tekrarlayan Endoskopi Kohortu

Hakan Öztürk, Ayşe Can, Fatma Eren Kurtipek, Ödül Eğritaş Gürkan, Sinan Sarı, Buket Dalgıç

Division of Pediatric Gastroenterology, Hepatology, and Nutrition, Gazi University, Faculty of Medicine, Ankara, Türkiye

ABSTRACT

Objective: The pathogenesis of eosinophilic esophagitis (EoE) is not fully understood, food allergens are considered the primary triggers. Nevertheless, the frequent coexistence of EoE with atopic diseases suggests that aeroallergens may also play an important role in disease development and exacerbation. The aim of this study was to assess the relationship between seasonality and disease activity, as determined by esophageal eosinophil counts and endoscopic findings, in children with EoE.

Methods: This retrospective cohort study included pediatric patients diagnosed with EoE who were followed at our center between January 1, 2010, and June 1, 2026. Demographic, clinical, laboratory, endoscopic, and histopathological data were retrospectively collected from the medical records and analyzed. The spring and summer months were defined as the high-pollen season, whereas the autumn and winter months were defined as the low-pollen season.

Results: A total of 526 upper gastrointestinal endoscopic examinations were performed in the 130 patients during the study period. An accompanying allergic disease was present in 55 patients (42.3%), whereas 75 patients (57.7%) had no history of allergic disease. Asthma was the most common allergic comorbidity, affecting 33 patients (25.4%). Compared with the low-pollen season, the high-pollen season was associated with significantly higher median EREFS scores [2 (interquartile range [IQR], 0-4) vs. 1 (IQR, 0-3), p = 0.012] and significantly higher median peak histologic eosinophil counts [35 (IQR, 0-71) vs. 25 (IQR, 0-60)] eosinophils per high-power field, p = 0.013).

ÖZ

Amaç: Eozinofilik özofajitin (EoE) patogenezi tam olarak aydınlatılamamış olmakla birlikte, gıda alerjenleri hastalığın başlıca tetikleyicileri olarak kabul edilmektedir. Bununla birlikte, EoE'nin atopik hastalıklarla sıklıkla birlikte görülmesi, aeroalerjenlerin de hastalığın gelişimi ve alevlenmesinde önemli bir rol oynayabileceğini düşündürmektedir. Bu çalışmanın amacı, eozinofilik özofajitli çocuklarda özofagus eozinofil sayıları ve endoskopik bulgular ile belirlenen hastalık aktivitesi ile mevsimsellik arasındaki ilişkiyi değerlendirmektir.

Yöntemler: Bu retrospektif kohort çalışmasına, 1 Ocak 2010 ile 1 Haziran 2026 tarihleri arasında merkezimizde izlenen ve EoE tanısı alan pediatrik hastalar dahil edildi. Demografik, klinik, laboratuvar, endoskopik ve histopatolojik veriler tıbbi kayıtlardan geriye dönük olarak elde edilerek analiz edildi. İlkbahar ve yaz ayları yüksek polen sezonu, sonbahar ve kış ayları ise düşük polen sezonu olarak tanımlandı.

Bulgular: Çalışma dönemi boyunca 130 hastaya toplam 526 üst gastrointestinal sistem endoskopisi uygulandı. Hastaların 55'inde (%42,3) eşlik eden alerjik hastalık bulunurken, 75'inde (%57,7) alerjik hastalık öyküsü yoktu. Astım, 33 hastayı (%25,4) etkileyerek en sık görülen alerjik komorbiditeydi. Düşük polen sezonuyla karşılaştırıldığında, yüksek polen sezonunda medyan EREFS skorları [2 (çeyreklar arası aralık [IQR], 0-4) vs. 1 (IQR, 0-3), p = 0,012] ve medyan pik histolojik eozinofil sayıları [35 (IQR, 0-71) vs. 25 (IQR, 0-60) yüksek büyütmeli alandaki eozinofiller, p = 0,013] anlamlı olarak daha yüksekti.

Sonuç: Bu bulgular, aeroalerjenlerin EoE aktivitesinin mevsimsel modülasyonunda potansiyel bir rol oynadığını desteklemekte ve EoE'li

Cite this article as: Öztürk H, Can A, Eren Kurtipek F, Eğritaş Gürkan Ö, Sarı S, Dalgıç B. Seasonal variation in histologic and endoscopic activity of pediatric eosinophilic esophagitis: a repeated endoscopy cohort. Gazi Med J. 2026;37(4):551-554

Address for Correspondence/Yazışma Adresi: Hakan Öztürk, Division of Pediatric Gastroenterology, Hepatology, and Nutrition, Gazi University, Faculty of Medicine, Ankara, Türkiye
E-mail / E-posta: hakan.ozturk@gazi.edu.tr
ORCID ID: orcid.org/0000-0003-4871-6019

Received/Geliş Tarihi: 30.07.2026
Accepted/Kabul Tarihi: 21.09.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

Conclusion: These findings support a potential role for aeroallergens in the seasonal modulation of EoE activity and suggest that seasonal aeroallergen exposure should be considered during the clinical follow-up of pediatric patients with EoE.

Keywords: Eosinophilic esophagitis, pediatric, seasonality, aeroallergens, endoscopy

ÖZ

pediyatrik hastaların klinik izlemi sırasında mevsimsel aeroalerjen maruziyetinin dikkate alınması gerektiğini düşündürmektedir.

Anahtar Sözcükler: Eozinofilik özofajit, pediyatrik, mevsimsellik, aeroalerjenler, endoskopi

INTRODUCTION

Eosinophilic esophagitis (EoE) is a chronic, immune-mediated inflammatory disorder of the esophagus characterized by symptoms of esophageal dysfunction. In children, clinical presentation is age-dependent and may include feeding difficulties, vomiting, abdominal pain, dysphagia, food impaction, poor growth, and failure to thrive. According to the European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) guideline, the histological diagnosis of EoE is established by the presence of a peak eosinophil count of ≥15 eosinophils per high-power field (eos/HPF) in esophageal biopsy specimens in the appropriate clinical setting (1). Over the past two decades, both the incidence and prevalence of EoE have increased markedly, particularly in the pediatric population (2). Although the pathogenesis of EoE is not fully understood, food allergens are considered the primary triggers. Nevertheless, the frequent coexistence of EoE with atopic diseases, particularly allergic rhinitis and asthma, suggests that aeroallergens may also play an important role in disease development and exacerbation (3). Allergic comorbidities, including allergic rhinitis, asthma, and urticaria, are present in approximately 41% of patients with EoE (4). The aim of this study was to assess the relationship between seasonality and disease activity, as determined by esophageal eosinophil counts and endoscopic findings, in children with EoE.

MATERIALS AND METHODS

This retrospective cohort study included pediatric patients diagnosed with EoE who were followed at the Division of Pediatric Gastroenterology, Hepatology and Nutrition, Gazi University, Faculty of Medicine, between January 1, 2010, and June 1, 2026. Demographic characteristics, clinical features, laboratory findings (peripheral blood eosinophil percentages obtained within the one month preceding each endoscopic examination), endoscopic findings, and histopathological results were retrospectively extracted from the medical records and analyzed. Because patients underwent multiple follow-up endoscopic examinations during the study period, all endoscopic assessments were included in the analysis and categorized according to the season in which they were performed, regardless of treatment status. The diagnosis of EoE was made according to the ESPGHAN guideline after excluding secondary causes of esophageal eosinophilia, including gastroesophageal reflux disease, Crohn's disease, vasculitis, infections, and connective tissue disorders, and demonstrating a peak eosinophil count of ≥15 eos/HPF in esophageal biopsy specimens obtained during upper gastrointestinal endoscopy (1). Biopsy specimens obtained from the proximal, middle, and distal esophagus were reviewed. The highest eosinophil count identified in any biopsy specimen was recorded as the peak eosinophil count. Endoscopic findings were assessed using the Eosinophilic Esophagitis Endoscopic Reference Score (EREFS), which evaluates edema, rings, exudates, furrows, and strictures (5). Endoscopic examinations were categorized according to the season in which they were performed: Winter (December-February), spring (March-May), summer (June-August), and autumn (September-November). To evaluate the potential effect of aeroallergen exposure, seasons were dichotomized into a high-pollen season (spring/summer) and a low-pollen season (autumn/winter), reflecting the typical pattern of airborne pollen exposure in Türkiye (6-8). The study protocol was approved by the Ethics Committee of Gazi University (approval number: 1450; date: July 28, 2026).

Statistical Analysis

Data analysis was undertaken using IBM SPSS Statistics for Windows (version 22.0; IBM Corp., Armonk, NY, USA). The normality of continuous variables was evaluated with the Shapiro-Wilk test before statistical testing. Normally distributed data are presented as mean ± standard deviation, whereas skewed data are summarized as median and interquartile range (IQR). Categorical variables are reported as frequencies with corresponding percentages. Comparisons between the high- and low-pollen seasons were performed using the independent-samples t-test or the Mann-Whitney U test, according to data distribution. Associations involving categorical variables were examined using Pearson's chi-square test or Fisher's exact test when appropriate. Statistical significance was defined as a two-sided p-value below 0.05.

RESULTS

A total of 130 pediatric patients with EoE were included in the study. Of these, 96 (73.8%) were male and 34 (26.2%) were female. A total of 526 upper gastrointestinal endoscopic examinations were performed in the 130 patients during the study period. The mean age at the first endoscopic examination was 7.5 ± 4.4 years. An accompanying allergic disease was present in 55 patients (42.3%), whereas 75 patients (57.7%) had no history of allergic disease. Asthma was the most common allergic comorbidity, affecting 33 patients (25.4%), followed by food allergy in 12 (9.2%), allergic rhinitis in 7 (5.4%), and urticaria in 3 (2.3%). At the time of endoscopy, 50.5% of the patients were asymptomatic. The most common symptoms were food impaction (16.5%), vomiting (11.6%), dysphagia (9.1%), abdominal pain (4.3%), prolonged chewing/selective eating (4.2%), and drinking lots of water with meals (3.3%). The monthly distribution of the mean endoscopic EREFS scores and mean peak histologic eosinophil counts is shown in Figure 1.

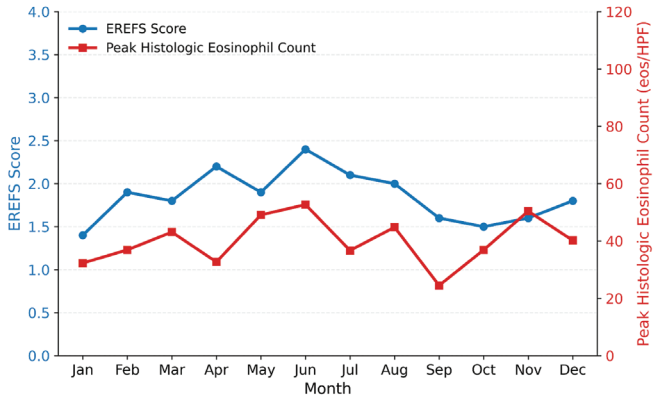

Figure 1. Monthly distribution of the mean endoscopic EREFS scores and mean peak histologic eosinophil counts.

EREFS: Eosinophilic Esophagitis Endoscopic Reference Score.

Both parameters demonstrated a similar temporal pattern, with higher values during the spring and summer months and lower values during autumn. When EREFS scores and peak histologic eosinophil counts were compared across the four seasons, both parameters were higher during spring and summer than during autumn and winter; however, these differences did not reach statistical significance (p = 0.071 and p = 0.086, respectively). In a prespecified secondary analysis, spring and summer were grouped as the high-pollen season, whereas autumn and winter were grouped as the low-pollen season. Compared with the low-pollen season, the high-pollen season was associated with significantly higher median EREFS scores [2 (IQR, 0-4) vs. 1 (IQR, 0-3), p = 0.012] and significantly higher median peak histologic eosinophil counts [35 (IQR, 0-71) vs. 25 (IQR, 0-60) eos/HPF, p = 0.013]. When the findings from patients' initial endoscopies were compared between the high- and low-pollen seasons, no significant differences were observed in EREFS scores or peak eosinophil counts (p = 0.600 and p = 0.847, respectively). However, when follow-up endoscopies were compared after excluding the patients' initial endoscopies, significant differences were observed in both EREFS scores and peak eosinophil counts (p = 0.005 and p = 0.002, respectively). Similarly, the median serum eosinophil percentage was significantly higher during the high-pollen season than during the low-pollen season [4.6% (IQR, 2.3-6.5) vs. 3.8% (IQR, 2.0-6.3), p = 0.042]. The odds of active histologic disease (peak eosinophil count ≥15 eos/HPF) were significantly higher during the high-pollen season than during the low-pollen season (odds ratio = 1.66, 95% confidence interval = 1.15-2.38; p = 0.006).

DISCUSSION

In the present study, we demonstrated that EoE activity was significantly greater during the high-pollen season (spring and summer) than during the low-pollen season. Patients evaluated during the high-pollen season had higher endoscopic disease severity, higher peak histologic eosinophil counts. Although previous studies have investigated the seasonal variation of EoE, our study differs from the existing literature in several important aspects. Unlike many earlier reports focusing solely on newly diagnosed patients or on individuals presenting with food impaction, our cohort included both newly diagnosed and follow-up patients, including those receiving treatment whose symptoms had improved (9). In our study, no significant seasonal differences in EREFS scores or peak eosinophil counts were observed at the initial endoscopy, whereas significant differences emerged in follow-up endoscopies. This discrepancy may partly be explained by the variable interval between symptom onset and diagnosis, as patients may undergo their initial endoscopy at different stages of disease activity. In addition, follow-up endoscopies may be more frequently performed in response to treatment assessment, recurrence of symptoms, or other clinical indications, potentially resulting in a greater representation of periods of active disease and thereby contributing to the observed seasonal differences. Furthermore, all data were derived from a single tertiary referral center with standardized endoscopic, histopathological, and clinical evaluation, minimizing concerns regarding heterogeneity that may arise in multicenter or nationwide database studies.

The etiology and pathogenesis of EoE have not yet been fully elucidated. Although EoE is primarily considered a food antigen-driven disease, its frequent coexistence with other atopic conditions raises the possibility that aeroallergens may also contribute to disease activity (10). A distinct seasonal pattern has been demonstrated for asthma exacerbations, with Harju et al. (11) reporting significant seasonal variation in asthma-related hospitalizations. Similarly, seasonal allergic rhinitis is well known to peak between April and June, corresponding to periods of increased airborne pollen exposure (12). In Türkiye, although regional variations exist, airborne pollen concentrations are highest during the spring and summer months (6). Considering these observations, the increased EoE activity observed during the high-pollen season in our study supports the hypothesis that aeroallergen exposure may contribute to the etiopathogenesis and seasonal exacerbation of EoE.

Several studies have previously investigated the seasonal distribution of EoE. In one of the largest published cohorts, Wang et al. (9) evaluated 234 patients with newly diagnosed EoE and found that the number of diagnoses was significantly lower during the winter months than during the other seasons. Similarly, Watanabe et al. (13) reported that EoE diagnoses were more frequent during spring and summer and less common during winter. In a study conducted in the United States including 79 children with EoE, 45% of patients were diagnosed during the spring months (14). In addition, several case reports have described seasonal exacerbations of EoE during the pollen season in patients with concomitant allergic rhinitis or asthma (15,16). Our findings are consistent with these previous observations, demonstrating increased EoE activity during the high-pollen season. From a clinical perspective, these findings suggest that, particularly in patients with concomitant allergic diseases such as asthma or allergic rhinitis, withdrawal or step-down of maintenance therapy may be more appropriately considered after the high-pollen season rather than during periods of peak aeroallergen exposure. However, prospective studies are needed to determine whether such a strategy improves long-term disease control.

Study Limitations

This study has several limitations. First, its retrospective design may have resulted in incomplete clinical documentation and limited the availability of potentially relevant variables, such as detailed environmental exposure data and individual pollen sensitization profiles. Second, although our institution is located in Ankara and most of our patients reside in Ankara, it is a tertiary referral center that receives patients from many regions of Türkiye. Therefore, regional differences in pollen calendars and aeroallergen exposure may have influenced the observed seasonal patterns.Because the analyses were based on the timing of endoscopy rather than patients' geographic residence or region-specific pollen counts, regional variability in aeroallergen exposure could not be accounted for. Third, the severity and activity of concomitant allergic diseases, including asthma and allergic rhinitis, were not systematically documented at the time of endoscopy. Consequently, the temporal relationship between seasonal exacerbations of these allergic diseases and EoE activity could not be directly evaluated. Finally, due to incomplete treatment-related data in the medical records, we were unable to perform a detailed comparison of seasonal disease activity according to specific treatment modalities, such as elimination diets, proton pump inhibitors, or topical corticosteroids.

CONCLUSION

In conclusion, our study demonstrated that EoE activity was significantly higher during the high-pollen season in a cohort including both newly diagnosed and follow-up patients. Increased endoscopic severity, histologic eosinophilic inflammation, peripheral eosinophilia, and a higher prevalence of active histologic disease during spring and summer support a potential role for aeroallergens in the seasonal modulation of EoE activity. These findings suggest that seasonal aeroallergen exposure should be considered during the clinical follow-up of patients with EoE, particularly in those with concomitant atopic diseases such as asthma or allergic rhinitis. Prospective multicenter studies incorporating regional pollen monitoring and allergen sensitization data are warranted to confirm these observations and to determine whether seasonally tailored monitoring and treatment strategies can improve disease outcomes.

Ethics

Ethics Committee Approval: The study protocol was approved by the Ethics Committee of Gazi University (approval number: 1450; date: July 28, 2026).

Informed Consent: Informed consent was not required as this was a retrospective study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: H.Ö., Ö.E.G., S.S., B.D., Concept: H.Ö., B.D., Design: H.Ö., S.S., Data Collection or Processing: A.C., F.E.K., Analysis or Interpretation: H.Ö., S.S., Literature Search: H.Ö., Writing: H.Ö., S.S.

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

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

REFERENCES

1. Amil-Dias J, Oliva S, Papadopoulou A, Thomson M, Gutierrez-Junquera C, Kalach N, et al. Diagnosis and management of eosinophilic esophagitis in children: an update from the European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN). J Pediatr Gastroenterol Nutr. 2024; 79: 394-437.
2. Hruz P, Straumann A, Bussmann C, Heer P, Simon HU, Zwahlen M, et al. Escalating incidence of eosinophilic esophagitis: a 20-year prospective, population-based study in Olten County, Switzerland. J Allergy Clin Immunol. 2011; 128: 1349-50.e5.
3. González-Cervera J, Arias Á, Redondo-González O, Cano-Mollinedo MM, Terreehorst I, Lucendo AJ. Association between atopic manifestations and eosinophilic esophagitis: a systematic review and meta-analysis. Ann Allergy Asthma Immunol. 582-90.e2.
4. Thel HL, Anderson C, Xue AZ, Jensen ET, Dellon ES. Prevalence and costs of eosinophilic esophagitis in the United States. Clin Gastroenterol Hepatol. 2025; 23: 272-80.e8.
5. Hirano I, Moy N, Heckman MG, Thomas CS, Gonsalves N, Achem SR. Endoscopic assessment of the oesophageal features of eosinophilic oesophagitis: validation of a novel classification and grading system. Gut. 2013; 62: 489-95.
6. Bülteni AAP. Polen takvimi, 2026. Erişim adresi: https://polenalerji.ankara.edu.tr/polen-takvimi.
7. Yildiz YK, Yaytokgil SB, Vezir E. Evaluation of clinical and demographic characteristics in children with aeroallergen sensitization: a study based on skin prick test results. Allergol Immunopathol (Madr). 2025; 53: 46-54.
8. Bıçakçı A, Çelenk S, Altunoğlu MK, Bilişik A, Canıtez Y, Malyer H, et al. Allergenic airborne *Gramineae* (Grass) pollen concentrations in Turkey. Asthma Allergy Immunol. 2009; 7: 90-9.
9. Wang FY, Gupta SK, Fitzgerald JF. Is there a seasonal variation in the incidence or intensity of allergic eosinophilic esophagitis in newly diagnosed children? J Clin Gastroenterol. 2007; 41: 451-3.
10. Sekar S, Kelly K, Sheth D, Papademetriou M. Inhalation and inflammation: examining aeroallergens and eosinophilic esophagitis. Front Allergy. 2025; 6: 1609120.
11. Harju T, Keistinen T, Tuuponen T, Kivela SL. Seasonal variation in childhood asthma hospitalisations in Finland, 1972-1992. Eur J Pediatr. 1997; 156: 436-9.
12. Ecevit MC, Ozcan M, Haberal Can I, Cadalli Tatar E, Ozer S, Esen E, et al. Turkish guideline for diagnosis and treatment of allergic rhinitis (ART). Turk Arch Otorhinolaryngol. 2021; 59(Suppl 1): 1-157.
13. Watanabe A, Tanaka F, Sawada A, Nishida Y, Maruyama H, Ominami M, et al. Association between serum alder-specific immunoglobulin E positivity and seasonal onset of eosinophilic esophagitis. JGH Open. 2024; 8: e70052.
14. Suryawala K, Palle S, Altaf MA. Epidemiology, clinical presentation, and seasonal variation in the diagnosis of children with eosinophilic esophagitis in Oklahoma. South Med J. 2020; 113: 37-41.
15. Fogg MI, Ruchelli E, Spergel JM. Pollen and eosinophilic esophagitis. J Allergy Clin Immunol. 2003; 112: 796-7.
16. Vande Velde S, Willekens J, De Bruyne R, De Bruyne P, Van Biervliet S. A case of paediatric seasonal eosinophilic oesophagitis. Acta Gastroenterol Belg. 2024; 87: 524-6.

Gazi Med J 2026;37(4):555-564

DOI: http://dx.doi.org/10.12996/gmj.2026.4791

Percutaneous Endoluminal Radiofrequency Ablation Combined with Biliary Stenting for Malignant Biliary Obstruction: A Retrospective Single-center Study

Malign Biliyer Obstrüksiyonda Perkütan Endolüminal Radyofrekans Ablasyon ile Birlikte Biliyer Stentleme: Retrospektif Tek Merkezli Çalışma

Fatih Öncü, Koray Akkan, Baran Önal, Erhan Ilgıt

Department of Radiology, Gazi University, Faculty of Medicine, Ankara, Türkiye

ABSTRACT

Objective: To assess the safety and clinical outcomes of combining percutaneous endoluminal radiofrequency ablation (RFA) with biliary stenting for malignant biliary obstruction.

Methods: We retrospectively evaluated 37 consecutive patients treated at a single center between 2013 and 2025 with percutaneous endoluminal RFA followed by metal stent placement for malignant biliary obstruction. Technical success, biochemical response, procedure-related complications, primary stent patency, recurrent biliary obstruction (RBO), and overall survival were evaluated. The 30-day biochemical analysis was restricted to patients who were alive at day 30 and had paired bilirubin measurements available. Kaplan-Meier analysis was used to estimate death-censored primary stent patency and overall survival, and an additional competing-risk analysis was performed for RBO with death without prior RBO treated as a competing event.

Results: Technical success was achieved in all patients (100%). Among the 24 patients who were alive at day 30 and had paired bilirubin measurements available, mean total bilirubin decreased from 7.41 ± 5.84 to 2.04 ± 1.46 mg/dL and mean direct bilirubin from 5.69 ± 4.46 to 1.57 ± 1.15 mg/dL (both $p < 0.001$). Modified CIRSE Grade 3a procedure-related complications occurred in three patients (8.1%), whereas no Grade 4-5 procedure-related complications occurred and there was no procedure-related mortality (Grade 6). RBO developed in three of the 35 patients included in the time-to-event analyses (8.6%). The competing-risk cumulative incidence of RBO was 3.5% at 90 and

ÖZ

Amaç: Malign biliyer obstrüksiyonu olan hastalarda perkütan endolüminal radyofrekans ablasyonun (RFA) biliyer stentleme ile birlikte kullanımının güvenlik ve klinik sonuçlarını değerlendirmektir.

Yöntemler: Bu retrospektif, tek merkezli çalışmaya 2013-2025 yılları arasında malign biliyer obstrüksiyon nedeniyle perkütan endolüminal RFA ve ardından metal stent yerleştirilmesi uygulanan ardışık 37 hasta dahil edildi. Teknik başarı, biyokimyasal yanıt, işleme bağlı komplikasyonlar, primer stent açıklığı, rekürren biliyer obstrüksiyon (RBO) ve genel sağkalım değerlendirildi. Otuzuncu gün biyokimyasal analizi, 30. günde hayatta olan ve eşleştirilmiş bilirubin ölçümleri bulunan hastalarla sınırlandırıldı. Ölümle sansürlenmiş primer stent açıklığı ve genel sağkalım Kaplan-Meier yöntemiyle hesaplandı; ayrıca, öncesinde RBO gelişmeden gerçekleşen ölüm rakip olay kabul edilerek RBO için ek bir competing-risk analizi yapıldı.

Bulgular: Tüm hastalarda teknik başarı sağlandı (%100). Otuzuncu günde hayatta olan ve eşleştirilmiş bilirubin ölçümleri bulunan 24 hastada ortalama total bilirubin düzeyi 7,41 ± 5,84 mg/dL'den 2,04 ± 1,46 mg/dL'ye, ortalama direkt bilirubin düzeyi ise 5,69 ± 4,46 mg/dL'den 1,57 ± 1,15 mg/dL'ye düştü (her ikisi için $p < 0,001$). Üç hastada (%8,1) modifiye CIRSE Grade 3a işleme bağlı komplikasyon gelişirken, modifiye CIRSE Grade 4-6 işleme bağlı komplikasyon veya işleme bağlı mortalite görülmedi. Zaman-bağımlı analizlere dahil edilen 35 hastanın üçünde (%8,6) RBO gelişti. Competing-risk analizinde RBO'nun kümülatif insidansı 90 ve 180. günlerde %3,5, 365. günde ise %8,9 olarak bulundu. Ölümle sansürlenmiş medyan primer stent açıklığı 367

Cite this article as: Öncü F, Akkan K, Önal B, Ilgıt E. Percutaneous endoluminal radiofrequency ablation combined with biliary stenting for malignant biliary obstruction: a retrospective single-center study. Gazi Med J. 2026;37(4):555-564

Address for Correspondence/Yazışma Adresi: Fatih Öncü, Department of Radiology, Gazi University, Faculty of Medicine, Ankara, Türkiye
E-mail / E-posta: oncufatih@gmail.com
ORCID ID: orcid.org/0000-0003-1673-7253

Received/Geliş Tarihi: 17.08.2026
Accepted/Kabul Tarihi: 20.09.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

180 days and 8.9% at 365 days. Median death-censored primary stent patency was 367 days [95% confidence interval (CI), 181 days to not estimable], and median overall survival was 95 days (95% CI, 77-112 days).

Conclusion: Percutaneous biliary drainage followed by endoluminal RFA and SEMS placement was technically feasible and had an acceptable safety profile, with biliary decompression observed among patients evaluable at day 30. The low cumulative incidence of RBO was encouraging, but the small number of events, substantial competing mortality, and lack of a control group limit interpretation and prevent assessment of the independent contribution of RFA.

Keywords: Radiofrequency ablation, stents, cholestasis, extrahepatic, biliary tract neoplasms, radiology, interventional, survival analysis

ÖZ

gün [%95 güven aralığı (GA), 181 gün-üst sınır hesaplanamadı], medyan genel sağkalım ise 95 gün (%95 GA, 77-112 gün) olarak bulundu.

Sonuç: Perkütan biliyer drenajı takiben endolüminal RFA ve SEMS yerleştirilmesi teknik olarak uygulanabilir olup kabul edilebilir bir güvenlik profiline sahipti; 30. günde değerlendirilebilen hastalarda biliyer dekompresyon gözlendi. RBO'nun düşük kümülatif insidansı umut vericiydi; ancak olay sayısının azlığı, ölümün önemli bir rakip olay olması ve kontrol grubunun bulunmaması, bulguların yorumlanmasını sınırlamakta ve RFA'nın bağımsız katkısının değerlendirilmesine olanak vermemektedir.

Anahtar Sözcükler: Radyofrekans ablasyon, kendiliğinden genişleyebilen metalik stentler, kolestaz, safra yolu neoplazmları, girişimsel radyoloji, sağkalım analizi

INTRODUCTION

Malignant biliary obstruction (MBO) can arise from primary biliary tumors or from primary or metastatic malignancies that involve or compress the bile ducts. It is a common and clinically important problem in advanced hepatobiliary and pancreatic malignancies, in which many patients are not candidates for curative surgical resection at the time of diagnosis (1).

Progressive MBO may lead to cholangitis, hepatic dysfunction, and deterioration in quality of life. For these patients, effective biliary decompression and maintenance of biliary patency are central to palliative management (1,2). For palliative biliary drainage, self-expandable metal stents (SEMS) are generally favored over plastic stents because their larger caliber provides more durable patency (2). Nevertheless, recurrent biliary obstruction (RBO) remains an important clinical problem and may result from tumor ingrowth or overgrowth, epithelial hyperplasia, sludge formation, and other causes of stent dysfunction (2).

Several adjunctive local therapies have been investigated to improve biliary drainage and prolong stent patency, including photodynamic therapy, intraluminal brachytherapy, and endobiliary radiofrequency ablation (RFA). Endobiliary RFA delivers controlled thermal energy to malignant tissue within the biliary lumen, producing coagulative necrosis and local tumor debulking. When combined with biliary stenting, RFA may contribute to restoration of the biliary lumen and delay RBO (3-7). Randomized trials and recent meta-analyses have suggested potential benefits in stent patency and/or overall survival (OS); however, these benefits have not been consistent across studies (3-7).

Endobiliary RFA may be delivered through an endoscopic route or by percutaneous transhepatic access. Despite growing experience with endobiliary RFA overall, published evidence on its percutaneous use remains more limited. Early feasibility studies established the technical applicability of percutaneous intraductal RFA, while more recent comparative studies have provided additional evidence regarding its safety and clinical outcomes when combined with SEMS placement (8-11).

At our institution, percutaneous endoluminal RFA combined with biliary stenting has been used as a palliative treatment for MBO in patients who were not candidates for curative surgical resection. We therefore aimed to evaluate the safety and clinical outcomes of this approach in our clinical practice.

MATERIALS AND METHODS

Study Design and Patient Selection

We retrospectively identified consecutive patients treated at our institution between 2013 and 2025 who had histopathologically confirmed MBO, were not candidates for curative surgical resection, and underwent percutaneous endoluminal RFA followed by SEMS placement. The final study cohort comprised 37 patients.

At our institution, percutaneous endoluminal RFA was not routinely added to SEMS placement in all patients with unresectable MBO, and there was no standardized institutional protocol for selecting patients for adjunctive RFA.

Patients who underwent endoluminal RFA for obstruction of a previously placed biliary stent without primary stent placement, including those treated with balloon dilatation alone or balloon dilatation with additional stent placement, were not included in the present analysis.

Reporting followed the strengthening the reporting of observational studies in epidemiology recommendations (12). The study protocol was reviewed and approved by the Gazi University, Faculty of Medicine Clinical Research Ethics Committee (approval no: 8, date: 08.05.2026). All study procedures were carried out in accordance with the principles of the Declaration of Helsinki. All patients had provided written consent for their interventional treatment. Because the analysis was retrospective, additional study-specific consent was waived.

Procedure

Local anesthesia was used for all procedures, with conscious sedation added when clinically or procedurally required. Before RFA, percutaneous biliary drainage was performed to achieve biliary decompression and establish transhepatic biliary access. The combined RFA and SEMS-placement procedure performed after initial biliary drainage was defined as the index procedure.

A Habib™ EndoHPB RFA catheter (EMcision Ltd, London, UK) connected to a compatible radiofrequency generator was advanced

across the malignant biliary stricture under fluoroscopic guidance. Radiofrequency energy was delivered at 10 W for 90 s per application. According to stricture length, the probe was activated at two treatment levels for lesions measuring ≤40 mm and at three treatment levels for lesions >40 mm. At each treatment level, two consecutive 90-s applications were routinely performed, with sufficient overlap between adjacent treatment levels to cover the entire malignant segment and adequately include its proximal and distal margins. Control cholangiography was then performed; if luminal improvement at any treated level was considered inadequate, one or two additional 90-s applications were delivered at that level. The RFA technique and the 10-W/90-s protocol used for the index procedure remained unchanged throughout the study period. Each 90-s period of energy delivery was considered one individual RFA application. For repeat RFA performed during follow-up for in-stent obstruction, a power setting of 12 W was used.

Following RFA, a metallic stent (Wallstent®, Boston Scientific, Natick, MA) was deployed across the treated stricture. Stents measuring 50, 70, or 90 mm in length were used, with length selected according to the longitudinal extent and anatomical location of the obstruction. A 10-mm stent diameter was used in patients not requiring bilateral hilar stenting. In the four patients with hilar obstruction, 8-mm-diameter stents were deployed bilaterally in a side-by-side configuration through separate right and left transhepatic accesses. The two stents were positioned across the hilar obstruction and deployed simultaneously to provide drainage of both hepatic ductal systems. Stent number and configuration were otherwise selected according to the extent and anatomical location of the obstruction.

A procedure was considered technically successful when endoluminal RFA was completed, the intended SEMS configuration covered the treated stricture by at least 1 cm at both margins, and final cholangiography confirmed unobstructed contrast passage into the duodenum. A representative procedural sequence is shown in Figure 1.

Laboratory Assessment and Follow-up

Serum total and direct bilirubin levels were recorded before biliary drainage, immediately before RFA and stent placement, and 30 days after treatment. Early biochemical response at 30 days was assessed only in patients who were alive at day 30 and had paired pre-ablation and 30-day bilirubin measurements available. Of the 13 patients not included in this analysis, 11 died before the 30-day assessment and 2 had no postprocedural follow-up data.

Clinical follow-up was based on available outpatient and hospital records, laboratory findings, and imaging examinations. Imaging was performed when clinically indicated, particularly in patients with suspected RBO.

Complications related to the index procedure were retrospectively classified according to the modified Cardiovascular and Interventional Radiological Society of Europe (CIRSE) classification system (13). Because RFA and SEMS placement were performed as components of the same treatment procedure, adverse events occurring after the intervention were not attributed specifically to either RFA or stent placement. Postprocedural abdominal pain responsive to analgesics was recorded separately and was not assigned a modified CIRSE grade. Adverse events occurring before the index procedure and attributable to the preceding percutaneous biliary drainage were recorded separately and were not considered procedure-related complications of the index intervention. Procedure-related mortality was defined as death attributable to the RFA and/or stent-placement procedure.

Stent Patency and OS

Primary stent patency was defined as the interval from SEMS placement to the first episode of RBO. RBO was defined as recurrence of obstructive symptoms accompanied by cholestatic laboratory abnormalities and/or imaging evidence of biliary obstruction requiring reintervention (14,15).

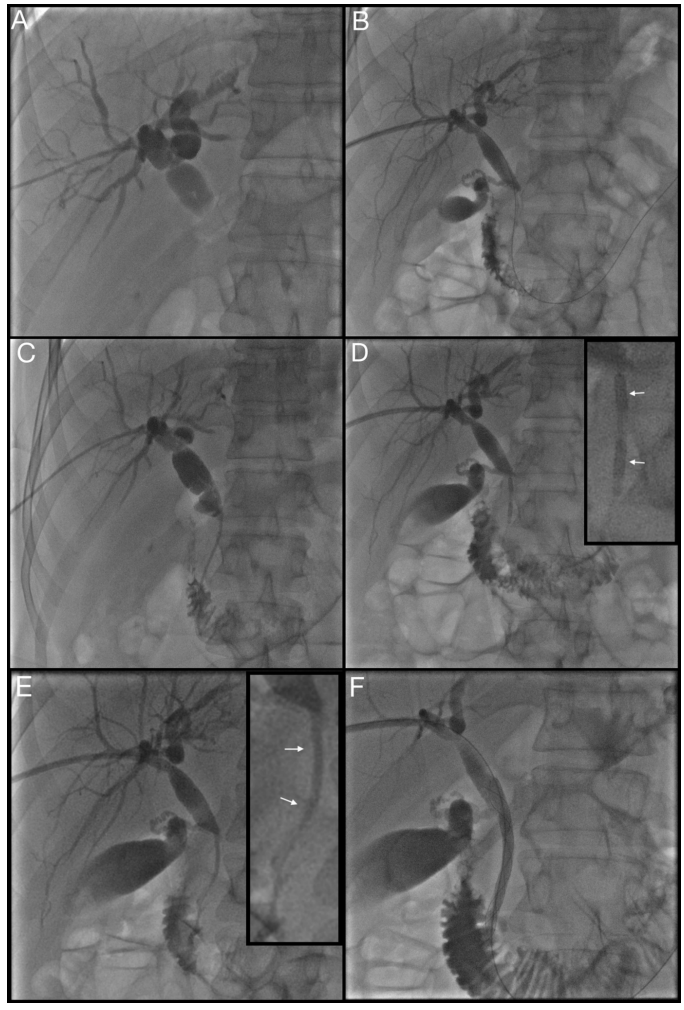

Figure 1. Representative fluoroscopic images of percutaneous endoluminal radiofrequency ablation and biliary stent placement for malignant biliary obstruction. (A) Percutaneous transhepatic cholangiography demonstrates the level of the common bile duct stricture. (B) The stricture is crossed with a guidewire. (C) A biliary drainage catheter is placed during the initial session. (D) During a subsequent session, the RFA catheter is positioned across the stricture and activated (white arrows). (E) Post-ablation cholangiography demonstrates increased luminal caliber at the treated segment before stent placement, consistent with local tumor debulking (white arrows). (F) Final cholangiography after stent placement demonstrates stent patency and free passage of contrast into the duodenum.

RFA: Radiofrequency ablation.

OS was calculated from the date of the index procedure to death from any cause or the last known follow-up. Patients for whom no postprocedural follow-up information was available were considered lost to follow-up and were excluded from the time-to-event analyses.

Statistical Analysis

Continuous data are reported as mean ± standard deviation and median (range), while categorical data are expressed as number and percentage.

The distribution of within-patient changes in bilirubin was examined before paired testing. As these differences were non-normally distributed, pre-ablation and day-30 bilirubin values were compared with the Wilcoxon signed-rank test. Statistical significance was set at a two-sided $p < 0.05$.

Kaplan-Meier curves were used for analyses of death-censored primary stent patency and OS. Patients who died without RBO were censored for the Kaplan-Meier patency analysis on the date of death (14,15). To account for death as a competing event, the cumulative incidence function (CIF) of RBO was additionally estimated using the Aalen-Johansen approach, with death without prior RBO treated as the competing event. Cumulative RBO incidence was estimated at predefined clinically relevant time points. Numbers at risk were displayed at 0, 100, 200, 300, 400, and 500 days below each time-to-event curve. Medians are presented with 95% confidence intervals (CIs) when calculable. Follow-up duration was summarized by the reverse Kaplan-Meier approach. Statistical analyses were performed using IBM SPSS Statistics, version 23.0 (IBM Corp., Armonk, NY, USA); the CIF was estimated using the survival package (version 3.7-0) in R software (version 4.4.1; R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Patient Characteristics

A total of 37 patients underwent percutaneous endoluminal RFA followed by SEMS placement. The mean age was 66.4 ± 11.7 years (median, 67 years; range, 48-90 years). Twenty-one patients (56.8%) were female and 16 (43.2%) were male.

Pancreatic carcinoma was the most common underlying malignancy (24/37, 64.9%), and the common bile duct was the most frequent level of obstruction (28/37, 75.7%).

The mean length of the malignant biliary stricture was 30.1 ± 10.5 mm, with a median of 29 mm (range, 16-59 mm). In lesions extending separately into the right and left hepatic ductal systems, the lengths of the involved segments were combined to represent the overall extent of the stricture. Detailed patient and lesion characteristics are summarized in Table 1.

Procedural Outcomes

Technical success was achieved in all patients (37/37, 100%). A total of 213 individual 90-s RFA applications were delivered during the 37 index procedures, corresponding to a mean of 5.8 applications per patient.

A total of 43 SEMS were deployed in 37 patients. Six patients received two SEMS, including all four patients with hilar obstruction and two additional patients in whom two stents were required because of stricture length and location.

The median interval between initial percutaneous biliary drainage and the index procedure was 9 days (range, 2-55 days). In 35 of 37 patients, RFA and SEMS placement were performed within 2-15 days after initial drainage. In one patient, the procedure was delayed to 32 days to allow resolution of a drainage-related subcapsular hepatic hematoma. In another patient, the interval was 55 days because of delayed return following an unplanned hospital departure after the initial drainage procedure. Procedural outcomes are summarized in Table 2.

Biochemical Outcomes

Both total and direct bilirubin levels had already fallen between initial biliary drainage and the pre-ablation assessment (Table 2).

Among the 24 patients who were alive at day 30 and had paired bilirubin measurements available, mean total bilirubin decreased from 7.41 ± 5.84 mg/dL before ablation to 2.04 ± 1.46 mg/dL at 30 days. The corresponding median values decreased from 5.48 mg/dL (range, 0.95-23.90 mg/dL) to 1.85 mg/dL (range, 0.75-7.91 mg/dL; $p < 0.001$). Mean direct bilirubin decreased from 5.69 ± 4.46 mg/dL to 1.57 ± 1.15 mg/dL, with corresponding median values decreasing from 4.13 mg/dL (range, 0.75-17.71 mg/dL) to 1.35 mg/dL (range, 0.57-6.22 mg/dL; $p < 0.001$).

The remaining 13 patients were not included in the paired biochemical analysis because 11 died before day 30 and 2 had no postprocedural follow-up data.

Complications

Postprocedural abdominal pain responsive to analgesic treatment occurred in 14 patients (37.8%).

Table 1. Patient and lesion characteristics.

Variable	Value
Total number of patients	37
Age, years, mean ± SD	66.4 ± 11.7
Age, years, median (range)	67 (48-90)
Sex, n (%)	
Female	21 (56.8)
Male	16 (43.2)
Underlying malignancy, n (%)	
Pancreatic carcinoma	24 (64.9)
Cholangiocarcinoma	6 (16.2)
Metastatic malignant biliary involvement	3 (8.1)
Gallbladder carcinoma	2 (5.4)
Gastric carcinoma	2 (5.4)
Level of biliary obstruction, n (%)	
Common bile duct	28 (75.7)
Common hepatic duct	5 (13.5)
Hilar	4 (10.8)
Stricture length, mm, mean ± SD	30.1 ± 10.5
Stricture length, mm, median (range)	29 (16-59)

Data are shown as mean ± standard deviation (SD), median (range), or n (%), according to variable type.

Table 2. Procedural, biochemical, and safety outcomes.

Variable	Value/mean ± SD	Median (range)	p-value
Procedural outcomes			
Technical success, n (%)	37/37 (100)		
Total SEMS deployed, n	43		
Patients receiving two SEMS, n (%)	6 (16.2)		
Bilateral SEMS placement for hilar obstruction, n (%)	4 (10.8)		
Individual 90-s RFA applications, total	213	-	-
RFA applications per patient, mean	5.8	-	-
Interval from biliary drainage to index RFA/SEMS, days	-	9 (2-55)	-
Biochemical parameters			
Before biliary drainage (n = 37)			
Total bilirubin, mg/dL	16.54 ± 5.91	—	
Direct bilirubin, mg/dL	12.76 ± 4.54	—	
Before RFA (n = 37)			
Total bilirubin, mg/dL	8.08 ± 5.17	—	
Direct bilirubin, mg/dL	6.21 ± 3.96	—	
Paired 30-day assessment (n = 24)			
Total bilirubin, mg/dL			<0.001
Pre-ablation	7.41 ± 5.84	5.48 (0.95-23.90)	
Day 30	2.04 ± 1.46	1.85 (0.75-7.91)	
Direct bilirubin, mg/dL			<0.001
Pre-ablation	5.69 ± 4.46	4.13 (0.75-17.71)	
Day 30	1.57 ± 1.15	1.35 (0.57-6.22)	
Safety and reintervention outcomes			
Postprocedural abdominal pain responsive to analgesics, n (%)	14 (37.8)		
Modified CIRSE Grade 3a procedure-related complications, n (%)	3 (8.1)		
Cholangitis	2 (5.4)		
Mild pancreatitis	1 (2.7)		
Modified CIRSE Grade 4-5 procedure-related complications	0		
Procedure-related deaths (modified CIRSE Grade 6)	0		
RBO requiring reintervention, n (%)	3/35 (8.6)		

Data are shown as mean ± standard deviation (SD), median (range), or n (%), as appropriate. Paired pre-ablation and day-30 biochemical values were analyzed with the Wilcoxon signed-rank test. CIRSE: Cardiovascular and Interventional Radiological Society of Europe, RBO: Recurrent biliary obstruction, RFA: Radiofrequency ablation, SEMS: Self-expandable metal stent.

According to the modified CIRSE classification, Grade 3a procedure-related complications occurred in three patients (8.1%), comprising cholangitis in two patients (5.4%) and mild pancreatitis in one patient (2.7%). Both patients with cholangitis recovered with intravenous antibiotic therapy, while the patient with pancreatitis recovered with supportive treatment.

No adverse event with findings suggestive of direct RFA-related thermal injury was identified. No modified CIRSE Grade 4-5 procedure-related complications occurred, and there was no procedure-related mortality (Grade 6).

One additional patient developed abdominal pain and a decrease in hemoglobin level after preprocedural biliary drainage and before undergoing RFA. Imaging demonstrated a subcapsular hepatic hematoma, and digital subtraction angiography showed no active arterial bleeding. The patient recovered with conservative management, and endoluminal RFA and stent placement were subsequently performed after resolution of the hematoma. Because this event occurred after the preceding biliary drainage procedure and before the index RFA and SEMS-placement procedure, it was classified as drainage-related rather than as a complication of the index intervention.

Follow-up, Primary Stent Patency, and OS

Two patients were lost to follow-up immediately after treatment and were therefore excluded from the time-to-event analyses. The remaining 35 patients were included in the analyses. The median

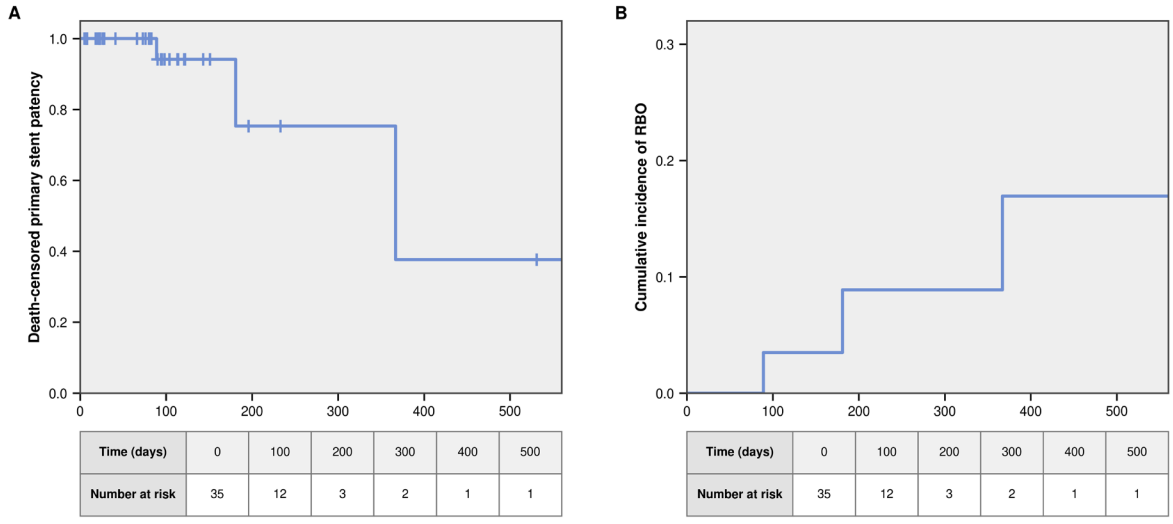

Figure 2. Time-to-event analysis of biliary stent outcomes. (A) Kaplan-Meier estimate of death-censored primary stent patency, with deaths occurring before RBO treated as censored observations. (B) Cumulative incidence of recurrent biliary obstruction estimated using the Aalen-Johansen approach, with death without prior RBO treated as a competing event. Numbers at risk at 0, 100, 200, 300, 400, and 500 days are shown below each panel.

RBO: Recurrent biliary obstruction.

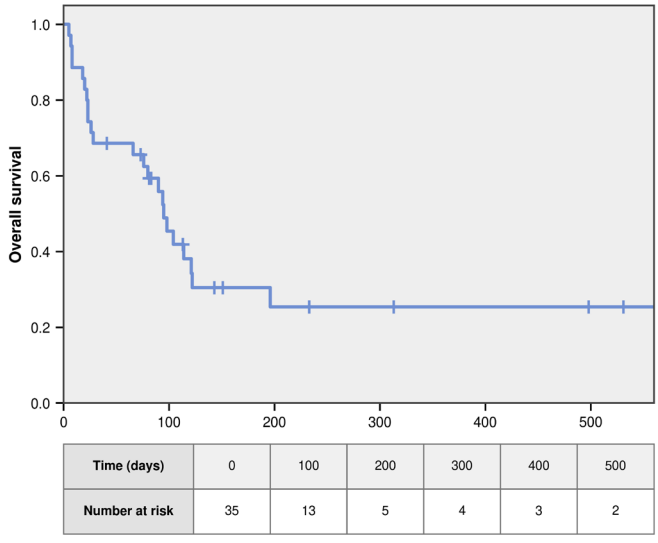

Figure 3. Kaplan-Meier curve showing overall survival after percutaneous endoluminal radiofrequency ablation and biliary stenting. Numbers at risk at 0, 100, 200, 300, 400, and 500 days are shown below the curve.

follow-up duration estimated using the reverse Kaplan-Meier method was 233 days (95% CI, 113-531 days).

RBO requiring reintervention occurred in three of the 35 patients (8.6%), at 89, 181, and 367 days after the index procedure.

In the first two patients, RBO was associated with stent dysfunction due to tumor ingrowth and was successfully managed with repeat endoluminal RFA at 12 W, followed by balloon dilatation and intraluminal sweeping. In the third patient, RBO was associated with stent dysfunction due to tumor ingrowth and overgrowth and was treated with repeat endoluminal RFA at 12 W, balloon dilatation and intraluminal sweeping, followed by additional SEMS placement, resulting in restoration of biliary patency.

Among the 35 patients, three experienced RBO, 23 died without a preceding RBO event, and nine were censored without RBO at their last available follow-up. In the competing-risk analysis, with death without prior RBO treated as a competing event, the cumulative incidence of RBO was 0% at 30 days, 3.5% at 90 days, 3.5% at 180 days, and 8.9% at 365 days.

In the conventional death-censored Kaplan-Meier analysis, median primary stent patency was 367 days (95% CI, 181 days to not estimable). The estimated death-censored primary stent patency rates at 30, 90, and 180 days were 100%, 94.1%, and 94.1%, respectively (Figure 2A). The competing-risk cumulative incidence curve is shown in Figure 2B. Numbers at risk at 0, 100, 200, 300, 400, and 500 days are displayed below both panels of Figure 2.

Among the 35 patients included in the OS analysis, 23 died during follow-up. The median OS was 95 days (95% CI, 77-112 days). Estimated cumulative survival rates at 30, 90, and 180 days were 68.6%, 55.9%, and 30.5%, respectively (Figure 3). Numbers at risk at the corresponding time points are displayed below the survival curve.

DISCUSSION

In this retrospective single-center series, percutaneous endoluminal RFA followed by SEMS placement was technically successful in all patients. Among patients evaluable at day 30, serum bilirubin levels decreased significantly. Modified CIRSE Grade 3a procedure-related complications occurred in three patients, whereas no modified CIRSE Grade 4-5 procedure-related complications occurred and there was no procedure-related mortality (Grade 6). Only three RBO events were documented during follow-up. The death-censored Kaplan-Meier median primary stent patency was 367 days, whereas competing-risk analysis yielded a cumulative RBO incidence of 8.9%

at 365 days. Median OS was 95 days. All three episodes of RBO were successfully managed with repeat RFA and adjunctive endobiliary treatment. Overall, these findings indicate that percutaneous endoluminal RFA combined with biliary stenting is technically feasible with an acceptable safety profile, but they also highlight that local biliary control and overall prognosis remain distinct in patients with advanced malignancy.

Our 100% technical success rate is in line with previous percutaneous RFA series (8-11,16-19). Cui et al. (17) reported a technical success rate of 98% in a retrospective series of 50 patients undergoing percutaneous intraductal RFA and SEMS placement, whereas other percutaneous series and comparative studies have similarly demonstrated high procedural success rates (18,19). Although available evidence does not establish a technical superiority of percutaneous over endoscopic RFA, direct transhepatic guidewire-based access may offer practical technical advantages in selected patients, particularly in proximal or anatomically complex biliary obstruction. Comparative studies of biliary drainage have reported high technical success with percutaneous access, with a potential advantage over endoscopic access particularly in high-level biliary obstruction (20-22). These studies address biliary drainage rather than RFA itself, so they cannot be taken as evidence that percutaneous RFA is superior to an endoscopic approach. Still, an established transhepatic tract may facilitate device advancement and sequential treatment of long or complex strictures in selected patients. Differences in catheter design, energy-delivery protocols, stent characteristics, obstruction level, and technical definitions across studies should be considered when comparing procedural outcomes.

By day 30, both total and direct bilirubin levels had fallen markedly. Mean total bilirubin decreased from 7.41 to 2.04 mg/dL, while mean direct bilirubin decreased from 5.69 to 1.57 mg/dL. These findings apply only to the 24 patients who were alive at day 30 and had paired laboratory measurements available and should not be extrapolated to the entire cohort. Importantly, bilirubin had already declined after the initial biliary drainage and before RFA. The median interval between drainage and RFA/SEMS placement was 9 days, with most procedures performed within 2-15 days; therefore, the pre-RFA bilirubin values reflect the effect of drainage over this interval. The subsequent biochemical improvement therefore cannot be attributed to RFA alone. RFA-induced coagulative necrosis and local tumor debulking may nevertheless help enlarge the effective biliary lumen before stent deployment and may subsequently contribute to maintenance of luminal caliber and biliary patency (9,23,24). Without a stent-only control group, however, the additional effect of RFA cannot be separated from that of drainage and stenting.

Primary stent patency is one of the most clinically relevant outcomes after endobiliary RFA because tumor ingrowth and overgrowth remain important causes of RBO after SEMS placement. Only three RBO events occurred in the present cohort, yielding a death-censored Kaplan-Meier median primary stent patency of 367 days. The estimated median primary stent patency was numerically longer than that reported in several previous percutaneous series. Cui et al. (17) reported a median stent patency of approximately 7 months in a single-center percutaneous series, while the more recent propensity score-matched study by Cui et al. (9) reported a median primary stent patency of 8.5 months after percutaneous RFA and stenting compared with 5.1 months after stenting alone. Other contemporary comparative studies and meta-analyses have reported mixed results regarding the effect of RFA on stent patency (18,25-28).

Because a substantial proportion of patients died before experiencing RBO, death represented an important competing event in the evaluation of biliary patency. The competing-risk analysis demonstrated cumulative RBO incidences of 3.5% at 90 and 180 days and 8.9% at 365 days. In contrast, the conventional Kaplan-Meier analysis treats deaths without RBO as censored observations and should therefore be interpreted specifically as death-censored primary stent patency. The divergence between the death-censored Kaplan-Meier estimate and the competing-risk cumulative incidence highlights the substantial impact of competing mortality on interpretation of stent outcomes in this advanced oncologic population.

The 367-day death-censored patency estimate should not, however, be overinterpreted. Only three RBO events occurred, and the upper limit of the 95% CI could not be estimated. Many patients died before a documented episode of RBO. The low number of RBO events is encouraging, but it also makes the late patency estimate imprecise. Direct comparisons with other studies are further complicated by differences in underlying malignancy, obstruction level, stent characteristics, definitions of RBO, follow-up, and the handling of death before RBO. Randomized trials and meta-analyses have likewise not shown a consistent effect of RFA on stent patency (3,4,7,27,28). Thus, the relatively long patency observed in some patients is encouraging, but our data cannot determine whether RFA provides additional benefit over SEMS placement alone.

The three episodes of RBO also provide practical information regarding reintervention. In each case, RBO was associated with stent dysfunction caused by tumor ingrowth, with additional tumor overgrowth in one patient. Repeat RFA at 12 W with balloon dilatation and intraluminal sweeping restored patency in the first two patients, while the third required an additional SEMS. This experience suggests that repeat intraductal RFA can be incorporated into the management of malignant stent dysfunction. Park et al. (29), in a multicenter propensity score-matched study, reported favorable time to RBO after in-stent RFA followed by uncovered SEMS placement for tumor ingrowth or overgrowth causing SEMS occlusion. Although our experience is limited to three cases, biliary drainage was restored in each, suggesting that repeat RFA may be a useful option when local tumor progression compromises an existing SEMS.

From a safety standpoint, the findings were also favorable. Abdominal pain responsive to analgesics was the most frequent postprocedural adverse event. The modified CIRSE Grade 3a procedure-related complications consisted of two cases of cholangitis and one case of mild pancreatitis, all of which resolved without permanent sequelae. Because RFA and SEMS placement were components of the same index intervention, these events could not be reliably attributed to either component individually. No adverse event with findings suggestive of direct RFA-related thermal injury was identified, and no modified CIRSE Grade 4-5 procedure-related complications occurred, and there was no procedure-related mortality (Grade 6). Previous studies and contemporary reviews have reported abdominal pain,

cholangitis, cholecystitis, pancreatitis, and, less frequently, biliary injury among adverse events associated with endobiliary RFA, although reported rates vary according to access route, device, ablation protocol, and complication definition (7,23,24). The subcapsular hepatic hematoma observed in one patient occurred after the preceding percutaneous biliary drainage procedure and before RFA and was accordingly considered a drainage-related event rather than a complication of the index RFA/SEMS intervention.

In contrast to the relatively long estimated stent patency, median OS was only 95 days and was shorter than that reported in several previous percutaneous RFA series (9,17-19). Survival in MBO is influenced by much more than local biliary control, including the biology and extent of the underlying malignancy, performance status, metastatic burden, and access to effective systemic treatment. The predominance of pancreatic carcinoma in our cohort may partly explain the difference from studies composed mainly of patients with cholangiocarcinoma. In the absence of a stent-only control group, our survival data also cannot be used to infer an effect of RFA on OS.

This difference between local control and survival makes patient selection particularly important. The benefit of improved local tumor control or prolonged stent patency is most likely to matter in patients expected to survive long enough to benefit from sustained biliary drainage. For patients with very limited life expectancy, adding an ablative procedure to stenting may offer less clinical value. This concept is supported by recent comparative data demonstrating an association between performance status and OS after treatment for MBO (9). Performance status, disease extent, anticipated survival, and the feasibility of subsequent systemic therapy may therefore all be relevant when considering RFA in addition to stenting. Future studies should focus not only on whether RFA improves biliary outcomes, but also on which patients are most likely to benefit from adding it to stenting.

The broader evidence on survival remains mixed. Several comparative studies and meta-analyses have suggested improved OS with RFA plus stenting (9,27,28), whereas individual randomized trials have reported variable results (3-5). A recent meta-analysis restricted to randomized trials suggested an OS benefit but did not demonstrate a significant improvement in stent patency (7). Differences in tumor type and location, RFA technique, stent choice, systemic treatment, and patient selection probably contribute to these inconsistent findings. Our survival results should therefore be viewed as descriptive of this cohort rather than as evidence for or against a survival benefit from RFA.

One practical strength of this series is the relatively uniform treatment strategy: All included patients underwent primary percutaneous endoluminal RFA followed by SEMS placement. Within the overall pathway of initial biliary drainage followed by RFA and SEMS placement, the procedure achieved high technical success, substantial biochemical improvement among evaluable patients, an acceptable safety profile, and relatively few documented episodes of RBO. Larger prospective, preferably multicenter, controlled studies are needed to determine the incremental value of percutaneous endoluminal RFA over stenting alone and to identify the patients most likely to benefit from durable local biliary control.

Study Limitations

This study has several limitations. Its retrospective, single-center design and small sample size introduce potential selection bias and limit generalizability. The heterogeneity of the underlying malignancies, with pancreatic carcinoma predominating, may also have influenced patency and survival outcomes. RFA was not applied according to a predefined selection algorithm, and its use varied during the study period, creating an additional potential source of treatment-selection bias. Without a contemporaneous stent-only control group, the independent contribution of RFA to biochemical response, primary stent patency, or survival cannot be determined. The 30-day biochemical analysis was limited to 24 surviving patients with paired measurements, while variation in the interval between biliary drainage and RFA/SEMS placement may have influenced bilirubin changes. Although death was accounted for as a competing event in an additional cumulative-incidence analysis, only three RBO events occurred, and the limited number of patients remaining at risk at later time points reduced the precision of both the competing-risk and death-censored Kaplan-Meier estimates. Finally, the study was not designed to identify predictors of early mortality or optimal patient-selection criteria, and changes in systemic oncologic and supportive care over the long study period may have influenced survival.

CONCLUSION

In conclusion, the overall percutaneous treatment strategy, including initial biliary drainage followed by endoluminal RFA and SEMS placement, was technically feasible and was associated with effective biliary decompression and an acceptable safety profile in patients with MBO who were not candidates for curative surgical resection. Because biliary drainage preceded RFA in all patients, the observed biochemical improvement cannot be attributed to RFA alone. However, the local coagulative necrosis and tumor-debulking effect of RFA may provide an adjunctive contribution to maintaining the biliary lumen and prolonging patency after initial decompression. The magnitude of this independent contribution cannot be established in the absence of a control group. The observed primary stent patency was encouraging, but the small number of RBO events and limited OS in this advanced oncologic population require cautious interpretation. Patient selection is likely to be important in determining the clinical value of this treatment. Prospective comparative studies should clarify both the added benefit of RFA and which patients are most likely to benefit from its use.

Ethics

Ethics Committee Approval: The study protocol was reviewed and approved by the Gazi University, Faculty of Medicine Clinical Research Ethics Committee (approval no: 8, date: 08.05.2026). All study procedures were carried out in accordance with the principles of the Declaration of Helsinki.

Informed Consent: All patients provided written consent for the procedure; separate consent for retrospective study participation was waived by the ethics committee.

Footnotes

Authorship Contributions

Surgical and Medical Practices: F.Ö., M.K.A., Concept: F.Ö., M.K.A., E.I., Design: F.Ö., B.Ö., E.I., Data Collection or Processing: F.Ö., B.Ö., Analysis or Interpretation: F.Ö., E.I., Literature Search: F.Ö., Writing: F.Ö.

Conflict of Interest: The authors report no conflicts of interest.

Financial Disclosure: This study received no external financial support.

REFERENCES

1. Rushbrook SM, Kendall TJ, Zen Y, Albazaz R, Manoharan P, Pereira SP, et al. British Society of Gastroenterology guidelines for the diagnosis and management of cholangiocarcinoma. Gut. 2024; 73: 16-46.
2. Dumonceau JM, Tringali A, Papanikolaou IS, Blero D, Mangiavillano B, Schmidt A, et al. Endoscopic biliary stenting: indications, choice of stents, and results: European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline—Updated October 2017. Endoscopy. 2018; 50: 910-30.
3. Albers D, Schmidt A, Schiemer M, Caca K, Wannhoff A, Sauer P, et al. Impact of endobiliary radiofrequency ablation on biliary drainage in patients with malignant biliary strictures treated with uncovered self-expandable metal stents: a randomized controlled multicenter trial. Gastrointest Endosc. 2022; 96: 970-9.
4. Jarosova J, Zarivnijova L, Cibulkova I, Mares J, Macinga P, Hujova A, et al. Endoluminal radiofrequency ablation in patients with malignant biliary obstruction: a randomised trial. Gut. 2023; 72: 2286-93.
5. Kang H, Han SY, Cho JH, Kim EJ, Kim DU, Yang JK, et al. Efficacy and safety of temperature-controlled intraductal radiofrequency ablation in advanced malignant hilar biliary obstruction: a pilot multicenter randomized comparative trial. J Hepatobiliary Pancreat Sci. 2022; 29: 469-78.
6. de Oliveira Veras M, de Moura DTH, McCarty TR, de Oliveira GHP, Gomes RSA, Landim DL, et al. Intraductal radiofrequency ablation plus biliary stent versus stent alone for malignant biliary obstruction: a systematic review and meta-analysis. Endosc Int Open. 2024; 12: E23-33.
7. Ramai D, Maida M, Smith ER, Wang Y, Spadaccini M, Previtera M, et al. Endoluminal radiofrequency ablation with stenting versus stenting alone in patients with malignant biliary obstruction: a meta-analysis of randomized trials. Endoscopy. 2025; 57: 272-81.
8. Kim KY, Yoon CJ, Lee JH, Lee CH, Hwang JH, Kim J. Percutaneous endobiliary radiofrequency ablation with stent placement in type IV hilar cholangiocarcinoma: a prospective comparison with stent placement alone. Eur J Radiol. 2024; 176: 111516.
9. Cui W, Huang JZ, Wang Q, Shi F, Gou Q, Chen XM, et al. Percutaneous endobiliary radiofrequency ablation and stent placement for unresectable malignant biliary obstruction: a propensity score matching retrospective study. BMC Gastroenterol. 2024; 24: 270.
10. Mizandari M, Pai M, Xi F, Valek V, Tomas A, Quaretti P, et al. Percutaneous intraductal radiofrequency ablation is a safe treatment for malignant biliary obstruction: feasibility and early results. Cardiovasc Intervent Radiol. 2013; 36: 814-9.
11. Wang Y, Cui W, Fan W, Zhang Y, Yao W, Huang K, et al. Percutaneous intraductal radiofrequency ablation in the management of unresectable Bismuth types III and IV hilar cholangiocarcinoma. Oncotarget. 2016; 7: 53911-20.
12. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007; 370: 1453-7.
13. Filippiadis D, Pereira PL, Hausegger KA, Ryan AG, Binkert CA. CIRSE standards of practice for the classification of complications: the modified CIRSE classification system. Cardiovasc Intervent Radiol. 2026; 49: 2-6.
14. Isayama H, Hamada T, Fujisawa T, Fukasawa M, Hara K, Irisawa A, et al. TOKYO criteria 2024 for the assessment of clinical outcomes of endoscopic biliary drainage. Dig Endosc. 2024; 36: 1195-210.
15. Isayama H, Hamada T, Yasuda I, Itoi T, Ryozawa S, Nakai Y, et al. TOKYO criteria 2014 for transpapillary biliary stenting. Dig Endosc. 2015; 27: 259-64.
16. Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. Control Clin Trials. 1996; 17: 343-6.
17. Cui W, Fan W, Lu M, Zhang Y, Yao W, Li J, et al. The safety and efficacy of percutaneous intraductal radiofrequency ablation in unresectable malignant biliary obstruction: a single-institution experience. BMC Cancer. 2017; 17: 288.
18. Wang J, Zhao L, Zhou C, Gao K, Huang Q, Wei B, et al. Percutaneous intraductal radiofrequency ablation combined with biliary stent placement for nonresectable malignant biliary obstruction improves stent patency but not survival. Medicine (Baltimore). 2016; 95: e3329.
19. Yu T, Zhang W, Li C, Wang C, Gong G, Wang L, et al. Percutaneous intraductal radiofrequency ablation combined with biliary stent placement for treatment of malignant biliary obstruction. Abdom Radiol (NY). 2020; 45: 3690-7.
20. Bian C, Fang Y, Xia J, Shi L, Huang H, Xiong Q, et al. Is percutaneous drainage better than endoscopic drainage in the management of patients with malignant obstructive jaundice? A meta-analysis of RCTs. Front Oncol. 2023; 13: 1105728.
21. Wang Y, Zhao X, She Y, Kang Q, Chen X. The clinical efficacy and safety of different biliary drainage in malignant obstructive jaundice: a meta-analysis. Front Oncol. 2024; 14: 1370383.
22. Angsuwatcharakon P, Kulpatcharapong S, Chuncharunee A, Khor C, Devereaux B, Moon JH, et al. The updated Asia-Pacific consensus statement on the role of endoscopic management in malignant hilar biliary obstruction. Endosc Int Open. 2024; 12: E1065-E1074.
23. Yamazaki H, Yamashita Y, Kitano M. Endobiliary radiofrequency ablation for malignant biliary obstruction. Ther Adv Gastroenterol. 2025; 18: 17562848251387191.
24. Inoue T, Yoneda M. Updated evidence on the clinical impact of endoscopic radiofrequency ablation in the treatment of malignant biliary obstruction. Dig Endosc. 2022; 34: 345-58.
25. Kong YL, Zhang HY, Liu CL, He XJ, Zhao G, Wang C, et al. Improving biliary stent patency for malignant obstructive jaundice using endobiliary radiofrequency ablation: experience in 150 patients. Surg Endosc. 2022; 36: 1789-98.
26. Oh D, Chong J, Song TJ, Park DH, Lee SS, Seo DW, et al. The usefulness of endobiliary radiofrequency ablation before metal stent placement in unresectable malignant hilar obstruction. J Gastroenterol Hepatol. 2022; 37: 2083-90.
27. Song S, Jin H, Cheng Q, Gong S, Lv K, Lei T, et al. Local palliative therapies for unresectable malignant biliary obstruction:

radiofrequency ablation combined with stent or biliary stent alone? An updated meta-analysis of nineteen trials. Surg Endosc. 2022; 36: 5559-70.

28. Tarar ZI, Farooq U, Gandhi M, Ghous G, Saleem S, Kamal F, et al. Effect of radiofrequency ablation in addition to biliary stent on overall survival and stent patency in malignant biliary obstruction: an updated systematic review and meta-analysis. Eur J Gastroenterol Hepatol. 2023; 35: 646-53.
29. Park N, Jung MK, Kim EJ, Paik WH, Cho JH. In-stent radiofrequency ablation with uncovered metal stent placement for tumor ingrowth/overgrowth causing self-expandable metal stent occlusion in distal malignant biliary obstruction: multicenter propensity score-matched study. Gastrointest Endosc. 2023; 97: 694-703.e2.

Gazi Med J 2026;37(4):565-571

DOI: http://dx.doi.org/10.12996/gmj.2026.4800

Association Between Regular Postpartum Breastfeeding Counseling and Maternal-infant Bonding: Second-month Postpartum Outcomes

Düzenli Postpartum Emzirme Danışmanlığı ile Anne-bebek Bağlanması Arasındaki İlişki: Postpartum İkinci Ay Sonuçları

Zahide Küçük[1], Aydan Biri[2]

[1]Zahide Küçük Gynecology and Obstetrics Clinic, Ankara, Türkiye
[2]Clinic of Obstetrics and Gynecology, Koru Hospital, Ankara, Türkiye

ABSTRACT

Objective: Maternal-infant bonding is central to postpartum psychological adaptation, and breastfeeding-related difficulties may adversely affect maternal emotional well-being and bonding. This study investigated the association between regular postpartum breastfeeding counseling and maternal-infant bonding during the second postpartum month.

Methods: In this single-center analytical observational study, 103 postpartum women were evaluated at 60 ± 7 days after delivery and grouped according to whether they received regular breastfeeding counseling during the first postpartum week. Maternal-infant bonding was assessed with the validated Turkish version of the Mother-to-infant bonding scale (MIBS). Groups were compared using Student's t-test and the chi-square test, and multivariable linear regression evaluated the independent association between regular counseling and total MIBS scores.

Results: Overall, 44 (42.7%) participants received regular counseling and 59 (57.3%) did not. Those who received counseling had significantly lower total MIBS scores (2.30 ± 2.15 vs. 6.39 ± 2.43; $p < 0.001$), with significant differences also across several exploratory item scores (all $p < 0.01$). In multivariable regression, regular counseling remained independently associated with lower total MIBS scores [$\beta = -4.51$, 95% confidence interval (CI): -5.43 to -3.58; $p < 0.001$], as did employment status ($\beta = 1.30$, 95% CI: 0.29-2.32; $p = 0.012$).

Conclusion: Regular postpartum breastfeeding counseling was associated with more favorable maternal-infant bonding during

ÖZ

Amaç: Anne-bebek bağlanması, postpartum dönemde psikolojik uyumun temel unsurlarından biridir. Emzirmeyle ilişkili güçlükler, annenin duygusal iyilik halini ve anne-bebek bağlanmasını olumsuz etkileyebilir. Bu çalışmada, düzenli postpartum emzirme danışmanlığı ile postpartum ikinci ayda anne-bebek bağlanması arasındaki ilişki araştırılmıştır.

Yöntemler: Tek merkezli, analitik ve gözlemsel nitelikteki bu çalışmada, doğum sonrası 60 ± 7. günde 103 kadın değerlendirilmiş ve postpartum ilk hafta boyunca düzenli emzirme danışmanlığı alıp almamalarına göre gruplandırılmıştır. Anne-bebek bağlanması, anne-bebek bağlanma ölçeğinin (MIBS) geçerliliği doğrulanmış Türkçe versiyonu kullanılarak değerlendirilmiştir. Gruplar Student's t-testi ve ki-kare testi ile karşılaştırılmış; düzenli danışmanlık ile toplam MIBS puanları arasındaki bağımsız ilişki çok değişkenli doğrusal regresyon analiziyle incelenmiştir.

Bulgular: Katılımcıların 44'ü (%42,7) düzenli emzirme danışmanlığı alırken, 59'u (%57,3) almamıştır. Danışmanlık alan grupta toplam MIBS puanları anlamlı düzeyde daha düşük bulunmuştur (2,30 ± 2,15'e karşı 6,39 ± 2,43; $p < 0,001$). Keşifsel olarak değerlendirilen bazı ölçek maddelerinin puanlarında da anlamlı farklılıklar saptanmıştır (tümünde $p<0,01$). Çok değişkenli regresyon analizinde, düzenli emzirme danışmanlığı daha düşük toplam MIBS puanlarıyla bağımsız olarak ilişkili bulunmuştur [$\beta = -4,51$; %95 güven aralığı (GA): -5,43 ile -3,58; $p < 0,001$]. Benzer şekilde, çalışma durumu da toplam MIBS puanlarıyla bağımsız olarak ilişkili bulunmuştur ($\beta = 1,30$; %95 GA: 0,29-2,32; $p = 0,012$).

Cite this article as: Küçük Z, Biri A. Association between regular postpartum breastfeeding counseling and maternal-infant bonding: second-month postpartum outcomes. Gazi Med J. 2026;37(4):565-571

Address for Correspondence/Yazışma Adresi: Zahide Küçük, Zahide Küçük Gynecology and Obstetrics Clinic, Ankara, Türkiye
E-mail / E-posta: drzahidekucuk@gmail.com
ORCID ID: orcid.org/0000-0003-4327-6958

Received/Geliş Tarihi: 02.09.2026
Accepted/Kabul Tarihi: 01.10.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

the second postpartum month. Given the observational design, this represents an association rather than a causal relationship, and randomized prospective studies are warranted to confirm it.

Keywords: Maternal-infant bonding, breastfeeding counseling, postpartum period, mother-to-infant bonding scale, maternal psychological adaptation

ÖZ

Sonuç: Düzenli postpartum emzirme danışmanlığı, postpartum ikinci ayda daha olumlu anne-bebek bağlanmasıyla ilişkili bulunmuştur. Çalışmanın gözlemsel tasarımı nedeniyle bu bulgu nedensel bir ilişkiden ziyade bir ilişkiyi göstermektedir. Bulguların doğrulanması için randomize prospektif çalışmalara ihtiyaç vardır.

Anahtar Sözcükler: Anne-bebek bağlanması, emzirme danışmanlığı, postpartum dönem, anne-bebek bağlanma ölçeği, maternal psikolojik uyum

INTRODUCTION

Maternal-infant bonding is a dynamic process that describes the emotional connection developed by a mother toward her infant. This bond begins during the prenatal period and continues throughout the child's life (1). A strong maternal-infant bond constitutes the foundation of the child's social, emotional, and cognitive development, facilitates maternal adaptation to the motherhood role, and contributes to the development of parenting skills (1,2). Positive bonding has also been shown to positively influence children's social skills during the preschool period (2). In contrast, impaired bonding has been associated with psychosocial problems, weakened parent-child relationships, and an increased risk of depression and anxiety later in life (3).

Maternal-infant bonding may be influenced by several biological, psychological, and social factors. Unplanned pregnancy, high parity, lack of social support, maternal stress, and postpartum depression have all been reported to negatively affect bonding levels (4-6). In addition, the childbirth process, postpartum pain severity, and maternal satisfaction with the birth experience may play important roles in postpartum psychological adjustment (7,8). Breastfeeding is considered one of the earliest and strongest forms of interaction between the mother and infant. However, difficulties experienced during breastfeeding may lead to stress, anxiety, and feelings of inadequacy in mothers, and mothers experiencing breastfeeding difficulties have been reported to exhibit lower bonding scores (3). Postpartum depression is also recognized as a major factor adversely affecting maternal-infant bonding (4-6).

Regular breastfeeding counseling may be associated with more positive maternal-infant bonding during the postpartum period by improving mothers' breastfeeding skills and self-confidence. Previous studies have demonstrated the beneficial effects of professional support and counseling on the childbirth experience (9). However, the number of studies investigating the association between regular breastfeeding counseling and maternal-infant bonding remains limited. To our knowledge, few studies have specifically evaluated this association using a validated maternal-infant bonding instrument in a routine postpartum care setting. Therefore, the present study aimed to compare maternal-infant bonding levels between mothers who received regular breastfeeding counseling during the postpartum period and those who did not.

MATERIALS AND METHODS

This single-center, analytical, observational study was conducted at the Zahide Küçük Gynecology and Obstetrics Clinic. Women receiving antenatal care at our institution who were eligible for follow-up were consecutively considered for enrollment. A total of 127 eligible pregnant women were initially enrolled. Twenty-four women were excluded from the final analysis: 8 delivered at another hospital because of relocation or other personal reasons, and 16 were excluded because of a history of stillbirth or neonatal death, a major congenital anomaly in the newborn, prolonged neonatal intensive care unit hospitalization (e.g., >7 days), severe maternal postpartum complications requiring intensive care, psychiatric illness, or diagnosed postpartum depression. Consequently, 103 postpartum women completed the study and were included in the final analysis (Figure 1).

All psychometric assessments were performed at a single time point during the second postpartum month. Following written informed consent, sociodemographic characteristics including maternal age, educational status, employment status, marital status, perceived income level, and smoking status were recorded. Obstetric and neonatal variables included parity, whether the pregnancy was planned, gestational age, mode of delivery, presence of episiotomy/perineal trauma, birth weight, infant sex, postpartum complications, and neonatal hospitalization history.

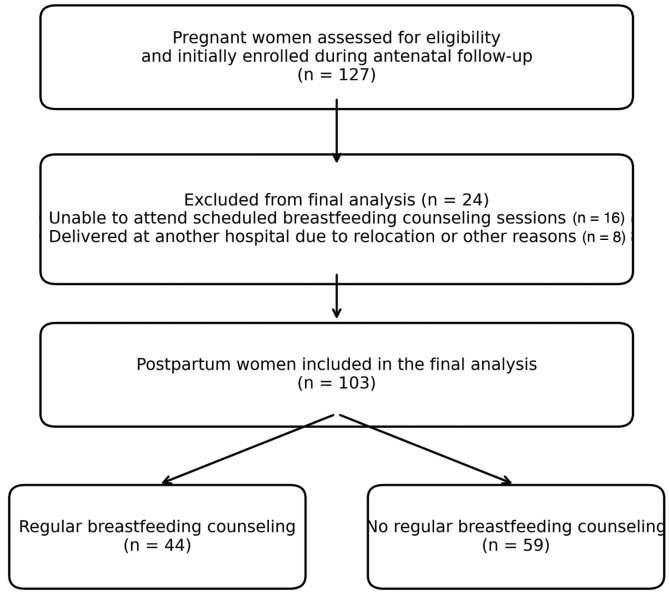

Figure 1. Participant flow diagram.

Maternal-infant bonding was assessed using the Turkish version of the mother-to-infant bonding scale (MIBS). The MIBS is an eight-item self-report questionnaire scored on a four-point Likert scale, with a total score range of 0-24 (10). The Turkish adaptation and psychometric validation of the scale were previously performed by Aydemir and Alparslan (11), demonstrating acceptable validity and reliability for assessing maternal-infant bonding in Turkish mothers. Lower total MIBS scores indicate more favorable maternal-infant bonding. In addition to the total MIBS score, individual MIBS item scores were evaluated descriptively.

Regular breastfeeding counseling consisted of a standardized postpartum breastfeeding education and support program delivered face-to-face by the same experienced breastfeeding nurse. Counseling was provided in three scheduled sessions during the first postpartum week, each lasting approximately 20 minutes. The educational content was standardized and included appropriate breastfeeding positioning, correct infant latch and holding techniques, breast care, prevention and management of common breastfeeding problems, and individualized responses to maternal questions and breastfeeding-related concerns.

Women aged 18 years or older who were able to read and understand Turkish, had undergone singleton live vaginal delivery, were within postpartum days 53-67 at the time of assessment, and provided written informed consent were included in the study. Women with a history of stillbirth or neonatal death, major congenital anomalies in the newborn, prolonged neonatal intensive care unit hospitalization (e.g., >7 days), severe maternal postpartum complications requiring intensive care, psychiatric illness, or diagnosed postpartum depression were excluded. Women with documented postpartum depression based on medical records or clinical psychiatric evaluation were excluded; however, no standardized screening instrument such as the Edinburgh postnatal depression scale was administered as part of the study protocol. In addition, women with active psychotic disorder, bipolar disorder, cognitive impairment interfering with reliable questionnaire completion, inability to establish effective communication, or refusal to complete the questionnaires were also excluded. All participant data were anonymized prior to analysis. No identifiable personal information was included in the study database, and all data were securely stored with access restricted exclusively to the research team.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics version 17.0 (IBM Corp., Armonk, NY, USA). Continuous variables were expressed as mean ± standard deviation, whereas categorical variables were presented as numbers and percentages. The distribution of continuous variables was assessed before analysis using visual inspection and normality testing. Comparisons between groups were performed using Student's t-test for continuous variables and the chi-square test for categorical variables, as appropriate. A multivariable linear regression model was constructed to evaluate the independent association between regular breastfeeding counseling and total MIBS score. Variables considered clinically relevant or showing significant between-group differences were included in the multivariable model. Regression model assumptions, including multicollinearity, residual distribution, linearity, and homoscedasticity, were evaluated. Analyses of individual MIBS items were considered exploratory; therefore, no formal adjustment for multiple comparisons was applied, and these findings should be interpreted cautiously. A post-hoc power analysis was performed using G*Power version 3.1.9.7 (Heinrich Heine University, Düsseldorf, Germany) for the primary outcome, namely the difference in total MIBS scores between the counseling and non-counseling groups. Based on the observed effect size (Cohen's d = 1.78), the final group sizes (44 and 59 participants), and a two-sided significance level of 0.05, the achieved statistical power exceeded 99.9%. Statistical significance was defined as a two-sided p-value <0.05.

Ethical Approval

The study was approved by the University of Health Sciences Türkiye, Ankara Bilkent City Hospital Clinical Research Ethics Committee (approval no: TABED 1-26-2584; date: 22 April 2026) and was conducted in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from all participants included in the study.

RESULTS

A total of 127 eligible pregnant women were initially enrolled in the study. After exclusion of 24 women (8 who delivered at another hospital and 16 who met one or more predefined exclusion criteria), 103 postpartum women were included in the final analysis. Participant selection and study flow are summarized in Figure 1.

Of the 103 participants, 44 (42.7%) received regular breastfeeding counseling, whereas 59 (57.3%) did not. The demographic and obstetric characteristics of the participants are presented in Table 1. No significant differences were observed between the groups regarding maternal age (31.84 ± 4.03 vs. 30.58 ± 4.02 years, respectively; p = 0.118) or gestational age at delivery (37.85 ± 1.86 vs. 38.44 ± 1.66 weeks; p = 0.344). Similarly, birth weights were comparable between the groups (3310 ± 250 g vs. 3231 ± 437 g; p = 0.296). The distribution of infant sex also did not differ significantly between the groups (p = 0.771). However, significant differences were identified with respect to educational status and employment status. Mothers who received regular breastfeeding counseling had higher educational attainment and were more frequently actively employed compared with those who did not receive counseling (p = 0.021 and p = 0.014, respectively). In contrast, perceived income level, smoking status, planned pregnancy, and neonatal hospitalization history were comparable between the groups (all p > 0.05). The most commonly reported postpartum complaints included breastfeeding-related problems, pain, and suture-related discomfort. Breastfeeding problems were reported significantly more frequently among mothers who did not receive regular breastfeeding counseling (p < 0.001).

Total MIBS scores and individual item scores are presented in Table 2. Mothers who did not receive regular breastfeeding counseling had significantly higher total MIBS scores compared with those who received counseling (6.39 ± 2.43 vs. 2.30 ± 2.15, respectively; p < 0.001).

Exploratory analysis of individual MIBS items demonstrated that mothers receiving regular breastfeeding counseling had significantly lower scores in items assessing loving feelings, anger, neutral feelings, joyfulness, disappointment, and irritability (all p < 0.01). In

Table 1. Comparison of demographic and clinical characteristics according to regular breastfeeding counseling status.

Variable	Regular counseling (n = 44)	No regular counseling (n = 59)	p
Age (years)	31.84 ± 4.03	30.58 ± 4.02	0.118
Parity	1.77 ± 0.80	2.20 ± 1.11	**0.024**
Gestational week	37.85 ± 1.86	38.44 ± 1.66	0.344
Birth weight (g)	3310 ± 250	3231 ± 437	0.296
Female infant sex	23 (52.3%)	33 (55.9%)	0.771
Smoking	4 (9.1%)	6 (10.2%)	0.385
Planned pregnancy	35 (79.5%)	45 (76.3%)	0.462
Neonatal hospitalization	8 (18.2%)	11 (18.6%)	0.892
Educational status			**0.021**
Primary school	1 (2.3%)	1 (1.7%)	
Secondary school	1 (2.3%)	8 (13.6%)	
High school	8 (18.2%)	18 (30.5%)	
University	25 (56.8%)	29 (49.2%)	
Master's degree	9 (20.5%)	3 (5.1%)	
Employment status			**0.014**
Unemployed	22 (50.0%)	47 (79.7%)	
Employed	21 (47.7%)	12 (20.3%)	
Other	1 (2.3%)	0 (0.0%)	
Income perception			0.512
Income lower than expenses	1 (2.3%)	3 (5.1%)	
Income equal to expenses	32 (72.7%)	42 (71.2%)	
Income higher than expenses	11 (25.0%)	14 (23.7%)	
Most common postpartum complaint			
Breastfeeding problems	14 (31.8%)	52 (88.1%)	**<0.001**
Pain	15 (34.1%)	6 (10.2%)	**<0.001**
Suture-related complaints	12 (27.3%)	0 (0.0%)	**<0.001**

Data are presented as mean ± standard deviation or n (%). Participants were allowed to select more than one postpartum complaint.

contrast, no significant between-group differences were observed for the items related to dislike and protectiveness.

In multivariable linear regression analysis (Table 3), regular breastfeeding counseling remained independently associated with lower total MIBS scores [β = -4.51, 95% confidence interval (CI): -5.43 to -3.58; p < 0.001].

Employment status was also independently associated with total MIBS scores (β = 1.30, 95% CI: 0.29-2.32; p = 0.012). In contrast, maternal age, parity, educational status, planned pregnancy, and smoking status were not significantly associated with total MIBS scores. Figure 2 presents the distribution of total MIBS scores according to study groups using boxplot analysis. Mothers who received regular breastfeeding counseling demonstrated lower median MIBS scores and a more homogeneous score distribution.

Table 2. Comparison of MIBS scores according to regular breastfeeding counseling status.

Score	No regular counseling	Regular counseling	p
Total MIBS score	6.39 ± 2.43	2.30 ± 2.15	**<0.001**
MIBS item 1 (Loving)	0.73 ± 0.52	0.11 ± 0.32	**<0.001**
MIBS item 2 (Angry)	0.97 ± 0.59	0.45 ± 0.50	**<0.001**
MIBS item 3 (Neutral)	0.73 ± 0.45	0.18 ± 0.45	**<0.001**
MIBS item 4 (Joyful)	0.69 ± 0.77	0.32 ± 0.52	**0.004**
MIBS item 5 (Dislike)	0.17 ± 0.50	0.05 ± 0.30	0.120
MIBS item 6 (Protective)	0.15 ± 0.48	0.14 ± 0.35	0.844
MIBS item 7 (Disappointment)	1.31 ± 0.68	0.41 ± 0.73	**<0.001**
MIBS item 8 (Irritable)	1.66 ± 0.82	0.66 ± 0.64	**<0.001**

Data are presented as mean ± standard deviation. Individual item analyses were exploratory. MIBS: Mother-to-infant bonding scale.

Table 3. Multivariable linear regression analysis of independent variables associated with total MIBS score.

Variable	B	95% CI	p
Regular breastfeeding counseling	-4.51	-5.43 to -3.58	**<0.001**
Age	0.03	-0.09 to 0.15	0.590
Parity	0.16	-0.38 to 0.69	0.567
Educational status	-0.12	-0.73 to 0.49	0.700
Employment status	1.30	0.29 to 2.32	**0.012**
Planned pregnancy	-0.46	-1.49 to 0.57	0.375
Smoking	0.19	-1.26 to 1.65	0.792

Model: R^2 = 0.527, adjusted R^2 = 0.492, model p < 0.001. CI: Confidence interval, MIBS: Mother-to-infant bonding scale.

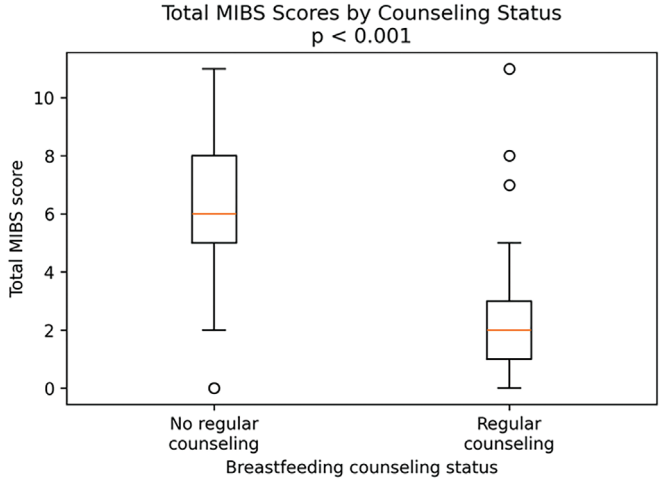

Figure 2. Boxplot distribution of total mother-to-infant bonding scale (MIBS) scores according to regular breastfeeding counseling status. Mothers who received regular breastfeeding counseling demonstrated lower median MIBS scores compared with mothers who did not receive counseling (p < 0.001). Lower MIBS scores indicate more favorable maternal-infant bonding.

DISCUSSION

The principal finding of the present study is that regular postpartum breastfeeding counseling was associated with significantly lower MIBS scores, suggesting more favorable maternal-infant bonding during the second postpartum month. Given that lower MIBS scores reflect stronger maternal bonding, these findings are consistent with more favorable maternal-infant bonding and suggest that regular breastfeeding counseling may be associated not only with breastfeeding practices but also with the emotional relationship established between the mother and her infant. However, given the observational design, these findings should be interpreted as an association rather than as evidence of a causal effect.

One of the methodological strengths of this study is the use of multivariable linear regression analysis. Even after adjustment for potential confounding variables including maternal age, parity, educational status, employment status, planned pregnancy, and smoking status, regular breastfeeding counseling remained independently associated with lower total MIBS scores. Nevertheless, the observational design also raises the possibility of reverse causation. Mothers with greater motivation, stronger early maternal attachment, or higher engagement with healthcare services may have been more likely to participate in regular breastfeeding counseling. Consequently, the observed association may partially reflect baseline participant characteristics rather than the counseling intervention alone.

Breastfeeding-related difficulties are known to contribute to maternal stress, anxiety, and feelings of inadequacy. The breastfeeding self-efficacy model proposed by Dennis (12) demonstrated that breastfeeding success is closely associated with maternal psychological confidence and emotional adjustment. Similarly, Qi et al. (6) reported that parenting self-efficacy mediates the relationship between postpartum depression and maternal bonding. In our cohort, breastfeeding-related problems were substantially more frequent among mothers who did not receive regular counseling (88.1% vs. 31.8%), supporting the hypothesis that breastfeeding counseling may be associated with benefits extending beyond technical education by enhancing maternal confidence and reducing emotional distress during the postpartum period.

Recent interventional studies have further demonstrated the beneficial effects of professional breastfeeding support on breastfeeding behaviors and maternal adaptation. In a randomized controlled trial conducted by Sönmez et al. (13), hybrid breastfeeding counseling significantly improved maternal breastfeeding behaviors and breastfeeding success. Our findings are consistent with these observations and additionally suggest that regular counseling may be associated with more favorable maternal-infant bonding. Although individual MIBS items were analyzed descriptively to better characterize maternal emotional responses, the scale was originally developed as a global measure of maternal-infant bonding rather than as a set of independent psychometric subdomains. Therefore, findings related to individual items should be regarded as exploratory and interpreted cautiously. Furthermore, the study by Daglar and Nur (14) demonstrated that social support, maternal emotional status, and postpartum psychological adjustment are closely associated with maternal-infant bonding, further supporting the psychosocial framework underlying our findings.

The influence of childbirth experience on postpartum psychological adaptation should also be considered. Severe postpartum pain and negative childbirth experiences have been associated with an increased risk of postpartum depression and impaired maternal emotional adjustment (7,8). In the Cochrane review by Bohren et al. (9), continuous professional support during childbirth was shown to improve maternal satisfaction and positively influence the overall birth experience. In this context, sustained breastfeeding counseling during the postpartum period may function as an important supportive mechanism by reducing emotional stress and facilitating maternal adaptation to infant care with greater confidence and emotional stability.

Another noteworthy finding of the present study was the independent association between employment status and total MIBS scores. Mothers who received regular breastfeeding counseling also had higher educational attainment and employment rates, suggesting that this group may initially have had a more advantageous profile with respect to healthcare access, health literacy, and postpartum self-management skills. Therefore, the observed association with maternal-infant bonding may not be attributable solely to breastfeeding counseling itself. Although several sociodemographic variables were adjusted for in the multivariable regression analysis, residual confounding and unmeasured psychosocial factors cannot be entirely excluded. Mothers receiving regular counseling may have differed from those who did not in terms of health literacy, healthcare utilization, motivation, family support, and psychosocial resources, all of which may influence both counseling participation and maternal-infant bonding outcomes.

Despite these limitations, structured breastfeeding counseling represents a practical and easily implementable component of routine postpartum care. Because breastfeeding counseling is already incorporated into many maternity care settings, optimizing and standardizing counseling delivery may represent a realistic strategy for strengthening postpartum support. If confirmed in prospective randomized studies, integrating standardized breastfeeding counseling into postpartum follow-up programs may contribute to maternal psychosocial well-being and maternal-infant bonding. Future multicenter prospective studies with randomized designs and comprehensive assessment of psychosocial and breastfeeding-related factors are warranted to clarify the causal relationship between breastfeeding counseling and maternal-infant bonding.

Study Limitations

First, because of its observational cross-sectional design, causality cannot be inferred, and the findings should be interpreted as demonstrating an association rather than a causal effect. Second, participation in the breastfeeding counseling program was not randomized, introducing the possibility of self-selection bias and reverse causation. Mothers receiving counseling had higher educational attainment and employment rates, which may reflect differences in health literacy, healthcare utilization, and psychosocial resources. Although multivariable regression analysis was performed, residual confounding cannot be completely excluded. Third, several potentially important psychosocial and breastfeeding-related variables, including breastfeeding exclusivity and duration, maternal anxiety, perceived social support, sleep quality, family

environment, breastfeeding motivation, infant temperament, maternal fatigue, and quality of the marital relationship, were not systematically collected and therefore could not be included in the multivariable analysis. Fourth, although women with clinically diagnosed postpartum depression were excluded, no standardized screening instrument such as the Edinburgh postnatal depression scale was administered; consequently, undiagnosed or subclinical depressive symptoms may have remained unrecognized (4-6,15). In addition, pregnancy-related anxiety and prenatal attachment, which may be affected by obstetric complications such as preeclampsia and gestational diabetes mellitus, were not assessed; these unmeasured factors could also influence maternal-infant bonding (16). Furthermore, although the counseling protocol was standardized with respect to counselor, educational content, session frequency, and duration, educational interventions are inherently influenced by participant-related factors, including educational attainment, health literacy, motivation, and psychosocial characteristics. Accordingly, individual variability in response to counseling cannot be completely eliminated. Finally, maternal-infant bonding was assessed using a self-report instrument without observer-based confirmation, and the single-center design may limit the generalizability of the findings.

CONCLUSION

Regular postpartum breastfeeding counseling was independently associated with more favorable maternal-infant bonding during the second postpartum month. Given the observational design of the present study, these findings should be interpreted as demonstrating an association rather than a causal relationship. Structured breastfeeding counseling may represent a practical component of routine postpartum care and could contribute to maternal well-being when incorporated into comprehensive postpartum support programs. However, prospective multicenter studies, ideally incorporating randomized designs and comprehensive assessment of psychosocial and breastfeeding-related factors, are needed to confirm these findings and clarify whether the observed association reflects a causal relationship.

Ethics

Ethics Committee Approval: The study was approved by the University of Health Sciences Türkiye, Ankara Bilkent City Hospital Clinical Research Ethics Committee (approval no: TABED 1-26-2584; date: 22 April 2026) and was conducted in accordance with the principles of the Declaration of Helsinki.

Informed Consent: Written informed consent was obtained from all participants included in the study.

Acknowledgments

For transparency, the authors note that an artificial intelligence-assisted language model (ChatGPT, OpenAI) was utilized to support text editing and language correction. This assistance was limited to linguistic refinement; all scientific content, critical analysis, and final editorial decisions were made exclusively by the authors.

Footnotes

Authorship Contributions

Concept: Z.K., A.B., Design: Z.K., A.B., Data Collection or Processing: Z.K., A.B., Analysis or Interpretation: Z.K., A.B., Literature Search: Z.K., A.B., Writing: Z.K., A.B.

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

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

REFERENCES

1. Hailemeskel HS, Kebede AB, Fetene MT, Dagnaw FT. Mother-infant bonding and its associated factors among mothers in the postpartum period, Northwest Ethiopia, 2021. Front Psychiatry. 2022; 13: 893505.
2. Joas J, Möhler E. Maternal bonding in early infancy predicts children's social competences in preschool age. Front Psychiatry. 2021; 12: 687535.
3. Figueiredo B, Costa R. Mother's stress, mood and emotional involvement with the infant: 3 months before and 3 months after childbirth. Arch Womens Ment Health. 2009; 12: 143-53.
4. Kasamatsu H, Tsuchida A, Matsumura K, Shimao M, Hamazaki K, Inadera H, et al. Understanding the relationship between postpartum depression one month and six months after delivery and mother-infant bonding failure one year after birth. Psychol Med. 2020; 50: 161-9.
5. Davies SM, Silverio SA, Christiansen P, Fallon V. Maternal-infant bonding and perceptions of infant temperament: the mediating role of maternal mental health. J Affect Disord. 2021; 282: 1323-9.
6. Qi W, Wei Z, Lv H, Zhao J, Hu Y, Wang Y, et al. Postpartum depression and maternal-infant bonding: the mediating role of mentalizing and parenting self-efficacy. BMC Pregnancy Childbirth. 2025; 25: 667.
7. Eisenach JC, Pan PH, Smiley R, Lavand'homme P, Landau R, Houle TT. Severity of acute pain after childbirth and persistent pain and postpartum depression. Pain. 2008; 140: 87-94.
8. Hodnett ED. Pain and women's satisfaction with the experience of childbirth: a systematic review. Am J Obstet Gynecol. 2002; 186: S160-72.
9. Bohren MA, Hofmeyr GJ, Sakala C, Fukuzawa RK, Cuthbert A. Continuous support for women during childbirth. Cochrane Database Syst Rev. 2017; 7: CD003766.
10. Taylor A, Atkins R, Kumar R, Adams D, Glover V. A new mother-to-infant bonding scale: links with early maternal mood. Arch Womens Ment Health. 2005; 8: 45-51.
11. Aydemir HK, Alparslan Ö. Adaptation of mother to infant bonding scale to the Turkish society: Aydın sample. J Contemp Med. 2016; 6: 188-99.
12. Dennis CL. Theoretical underpinnings of breastfeeding confidence: a self-efficacy framework. J Hum Lact. 1999; 15: 195-201.
13. Sönmez T, Gökyıldız Sürücü Ş, Avcıbay Vurgeç B, Vural G, Akca Ö, Baltacı O, et al. Effects of hybrid breastfeeding counseling on women's breastfeeding behaviors: a randomized controlled longitudinal study. BMC Pregnancy Childbirth. 2025; 25: 663.

14. Daglar G, Nur N. Level of mother-baby bonding and influencing factors during pregnancy and postpartum period. Psychiatr Danub. 2018; 30: 433-40.
15. Sanni SO, Adeoye IA, Bella-Awusah TT, Bello OO. Influence of postpartum depression on maternal-infant bonding and breastfeeding practices among mothers in Abeokuta, Ogun State. Discov Ment Health. 2024; 4: 46.
16. Yeşilçınar İ, Kıncı MF, Ünver HC, Sivaslıoğlu AA. Pregnancy-related anxiety and prenatal attachment in pregnant women with preeclampsia and/or gestational diabetes mellitus: a cross-sectional study. J Clin Obstet Gynecol. 2023; 33: 27-35.

Gazi Med J 2026;37(4):572-575

DOI: http://dx.doi.org/10.12996/gmj.2026.4568

Malignant Transformation to Mucinous Adenocarcinoma in Chronic Fistula-in-ano

Kronik Anal Fistülde Müsinöz Adenokarsinomaya Malign Dönüşüm

Farah Madiehah Abdul Hadi[1], Khairul Hazim Hamdan[1], Aswini Bala Krishnan[2], Firdaus Hayati[3]

[1]Department of Surgery, Selayang Hospital, Selayang, Batu Caves, Selangor, Malaysia
[2]Department of Pathology, Selayang Hospital, Selayang, Batu Caves, Selangor, Malaysia
[3]Department of Surgery, Faculty of Medicine and Health Sciences, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia

ABSTRACT

Malignant transformation in chronic fistula-in-ano (FIA) is rare, typically occurring due to long-term inflammation in untreated fistulas, though the exact cause remains unclear. These cancers are often aggressive, and their symptoms can resemble those of a benign fistula, leading to delayed diagnosis. We report a 71-year-old gentleman with supralevator FIA who developed perianal mucinous adenocarcinoma. The patient had suffered from a recurrent fistula for 40 years despite multiple surgeries and presented to our centre with a non-healing FIA. Magnetic resonance imaging (MRI) revealed a complex fistula with several openings, and histopathology confirmed adenocarcinoma with mucinous differentiation. Chronic FIA is a common but its rare potential to transform into perianal mucinous adenocarcinoma emphasizes the need for early detection and intervention. Although uncommon, malignant transformation should be considered in longstanding or recurrent complex FIA, particularly when symptoms persist despite previous treatment. Early imaging, such as MRI, and multiple biopsies are crucial for identifying malignancy and ensuring proper treatment, especially radical surgery.

Keywords: Anal fistula, anal gland, perianal gland neoplasm, mucinous adenocarcinoma, rectal fistula

ÖZ

Kronik anal fistülde (FIA) malign dönüşüm nadir görülmekte olup, genellikle tedavi edilmemiş fistüllerde uzun süreli enflamasyonla ilişkilendirilmekle birlikte, kesin etiyolojisi henüz aydınlatılamamıştır. Bu maligniteler sıklıkla agresif seyretmekte ve klinik bulgularının benign fistüllerle benzerlik göstermesi nedeniyle tanıda gecikmelere yol açabilmektedir. Bu çalışmada, supralevator anal fistülü bulunan ve perianal müsinöz adenokarsinom gelişen 71 yaşındaki bir erkek hastanın sunulması amaçlanmıştır. Hasta, birden fazla cerrahi girişime rağmen 40 yıldır tekrarlayan anal fistül öyküsü bulunması ve fistülün iyileşmemesi nedeniyle merkezimize başvurmuştur. Manyetik rezonans görüntüleme (MRG) sonucunda birden fazla dış açıklığı bulunan kompleks bir fistül saptanmış, histopatolojik inceleme ile müsinöz farklılaşma gösteren adenokarsinom tanısı doğrulanmıştır. FIA yaygın görülen bir klinik durum olmakla birlikte, nadiren perianal müsinöz adenokarsinomaya dönüşebilmesi, erken tanı ve tedavinin önemini ortaya koymaktadır. Özellikle uzun süredir mevcut olan veya tekrarlayan kompleks anal fistüllerde, önceki tedavilere rağmen semptomların devam etmesi durumunda malign dönüşüm olasılığı göz önünde bulundurulmalıdır. Malignitenin erken dönemde saptanması ve uygun tedavi stratejisinin belirlenmesi açısından MRG gibi görüntüleme yöntemlerinin erken dönemde uygulanması ve çoklu biyopsilerle histopatolojik değerlendirmenin yapılması büyük önem taşımaktadır. Bu yaklaşım, özellikle radikal cerrahi gerektiren olgularda tedavinin doğru biçimde planlanmasına katkı sağlayabilir.

Anahtar Sözcükler: Anal fistül, anal bez, perianal bez neoplazisi, müsinöz adenokarsinom, rektal fistül

Cite this article as: Hadi FMA, Hamdan KH, Krishnan AB, Hayati F. Malignant transformation to mucinous adenocarcinoma in chronic fistula-in-ano. Gazi Med J. 2026;37(4):572-575

Address for Correspondence/Yazışma Adresi: Firdaus Hayati, Department of Surgery, Faculty of Medicine and Health Sciences, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia
E-mail / E-posta: m_firdaus@ums.edu.my
ORCID ID: orcid.org/0000-0002-3757-9744

Received/Geliş Tarihi: 19.10.2025
Accepted/Kabul Tarihi: 20.07.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

INTRODUCTION

Perianal malignancies arising from abscesses or fistulas can either be adenocarcinomas or squamous cell carcinomas, with adenocarcinomas being more common (2.9-10%) (1). Among these, mucinous adenocarcinomas are the most frequently observed subtype (1). Mucinous adenocarcinomas are characterized by the secretion of extracellular mucin, which makes up more than 50% of the tumour volume (1). Perianal fistulas are a common problem, with around 90% caused by infected anal glands, often leading to perianal abscesses (2). Approximately 10% of perirectal fistulas are linked to other factors, such as Crohn's disease, trauma, human immunodeficiency virus, sexually transmitted infections, radiation therapy, or foreign bodies (2).

In rare cases, perianal fistulas and abscesses can undergo malignant transformation, potentially developing into carcinomas. These cancers are believed to result from chronic inflammation and irritation in long-standing fistulas, though the exact link between the fistula and carcinoma remains unclear (3). The tumours are usually aggressive, often mucinous, and frequently misdiagnosed early on (4). Symptoms can initially mimic benign fistula cases, and biopsies may not detect the carcinoma, leading to a delay in diagnosis. As a result, tumours are often found at more advanced stages, typically in deeper regions like the ischiorectal fossa and perineum (3). It can also be difficult to determine whether the cancer developed independently or because of the fistula (3).

Given the rarity of this condition and the lack of controlled trials, this article presents a case of perianal mucinous adenocarcinoma arising from an untreated anorectal fistula. We share our experience from a tertiary care institution and provide a review of the relevant literature to guide early diagnosis and management.

CASE PRESENTATION

A 71-year-old gentleman presented with a recurrent fistula-in-ano (FIA) that had persisted for 40 years. Despite multiple previous surgeries at different healthcare facilities, he continued to experience persistent purulent discharge. On perianal examination, several tender external openings were noted, draining pus and mucin (Figure 1). The tumour marker namely carcinoembryonic antigen was within normal range.

Colonoscopy revealed an irregular internal opening in the anal canal without evidence of ulceration or mucosal inflammation. The colonoscopy was advanced to the caecum. An incidental ascending colon polyp (Paris 1sp) was identified and removed via cold snare polypectomy. Histopathological examination of the polyp showed no dysplasia or malignancy. Magnetic resonance imaging (MRI) of the pelvis demonstrated a grade 5 complex perianal fistula (St. James's University Hospital classification) with supralevator extension and multiple bilateral ischioanal collections (Figure 2). The fistulous tracts demonstrated heterogeneous T2 hyperintensity with enhancing solid and papillary components on post-contrast sequences, raising suspicion for malignant transformation.

The patient subsequently underwent examination under anaesthesia, debridement, and seton insertion (Figure 3). Tissue obtained from the debridement was sent for histopathological analysis, which revealed adenocarcinoma with mucinous differentiation (Figure 4). Histopathological examination demonstrated low-grade, well-differentiated adenocarcinoma with abundant extracellular mucin comprising approximately 80% of the tumour volume, consistent with mucinous adenocarcinoma. The tumour was characterized by mucin pools containing villoglandular proliferations of neoplastic epithelial cells. A distinctive feature was the direct continuity between the neoplastic glandular epithelium and the perianal epidermis at the sinus openings, with abrupt transitions to squamous epithelium within epithelialized fistulous tracts, supporting the origin of the malignancy from a chronic FIA. Immunohistochemical staining showed CK7 negativity with positivity for CK20 and CDX2, supporting a colorectal-type adenocarcinoma phenotype.

He was clinically staged as cT4N0Mx according to the TNM staging system, consistent with locally advanced disease. A staging computed tomography (CT) scan was planned prior to multidisciplinary team discussion to determine the suitability of neoadjuvant therapy followed by radical surgery. However, the patient's condition deteriorated rapidly, and he unfortunately succumbed before completion of CT staging and further oncological planning. Written informed consent was obtained from the patient's relatives for their anonymised information to be published in this article.

DISCUSSION

Determining whether FIA preceded the development of malignancy or whether malignancy led to fistula formation can be clinically challenging. To help establish malignant transformation in chronic FIA, Rosser proposed three diagnostic criteria: The fistula should be present for at least 10 years prior to carcinoma development; the tumour identified in the rectum or anal canal should represent direct extension from carcinoma within the fistula; and the internal opening of the fistula should communicate with the anal canal rather than the tumour itself (4). Our patient fulfilled most of the reported criteria. However, the lesion was not directly visualised on colonoscopy and was identified only on MRI.

Figure 1. Perineal examination during initial visit showing several external openings, which were tender to the touch, draining pus and mucin.

Perianal mucinous adenocarcinoma arising from chronic anorectal fistulas is rare. The exact relationship between chronic fistulas and malignant transformation remains unclear, although long-standing inflammation is considered a contributing factor. These tumours often grow slowly within the ischioanal region and may not extend into the bowel lumen, which can delay diagnosis. Therefore, early detection requires a high index of suspicion, particularly in patients with long-standing or recurrent fistulas. Tissue biopsy is essential when malignancy is suspected, and pelvic MRI may help differentiate between inflammatory fistulous disease and malignant transformation (5).

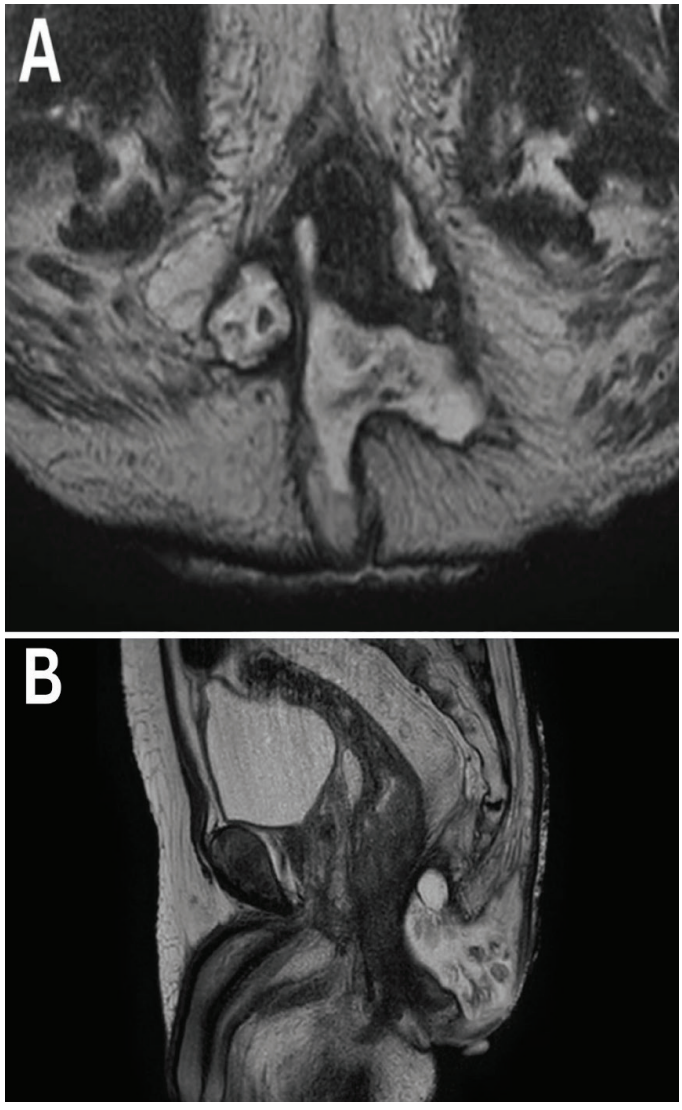

Figure 2. T2-weighted pelvic MRI demonstrating a grade 5 complex perianal fistula (St. James University Hospital classification) with supralevator extension and multiple bilateral ischioanal collections. The fistulous tracts communicate with the anal canal and extend across the sphincter complex with horseshoe configuration. The lesions show heterogeneous T2 hyperintensity with enhancing solid and papillary components on post-contrast sequences, with irregular thickening of the lower rectal wall, features suspicious for malignant transformation. Axial (A) and sagittal (B) views are shown.

MRI: Magnetic resonacne imaging.

Surgery remains the cornerstone of treatment, with adequate oncologic resection being critical for favourable outcomes. In many cases, radical procedures such as abdominoperineal resection are required to achieve complete tumour removal, as local excision alone is often insufficient. Neoadjuvant therapy may improve recurrence-free survival in selected patients; however, there is currently no consensus regarding the optimal use of neoadjuvant or adjuvant chemotherapy and radiotherapy, and treatment should be individualized. Early recognition and appropriate management are essential to prevent disease progression and improve prognosis (5).

Chronic and recurrent FIA should raise suspicion for malignant transformation, particularly in longstanding disease with persistent discharge or atypical features. Pelvic MRI and tissue biopsy play an important role in differentiating complex inflammatory fistulas from underlying mucinous adenocarcinoma, especially when colonoscopy findings are inconclusive. Early recognition and timely multidisciplinary management are essential, as delayed diagnosis often results in locally advanced disease requiring radical treatment.

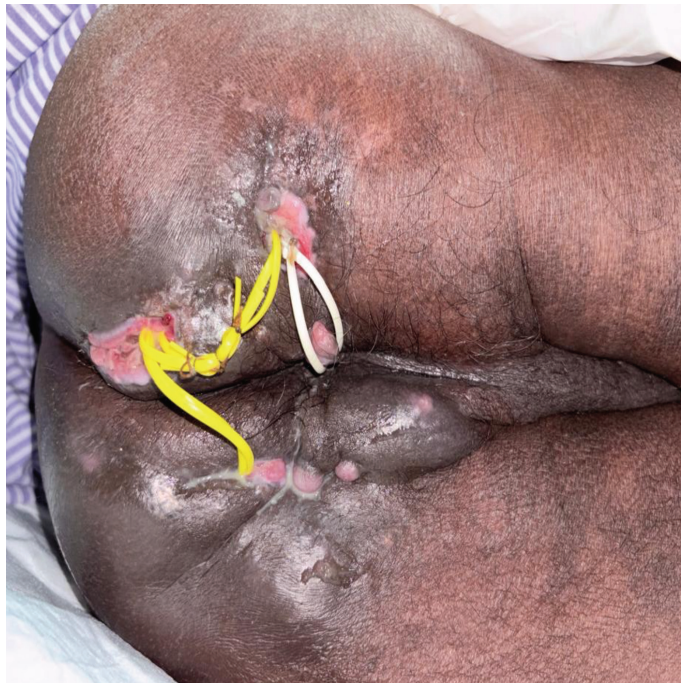

Figure 3. Perineal examination (post operative) following the examination under anesthesia and Hanley procedure, external openings were identified at 7 o'clock and 8 o'clock positions (3 cm from the anal verge), along with a supralevator internal opening at 8 o'clock (8 cm from the anal verge). Another external opening was located at 3 o'clock. The openings were interconnected at the 8 o'clock position. A mucin-filled cavity was present at all external openings.

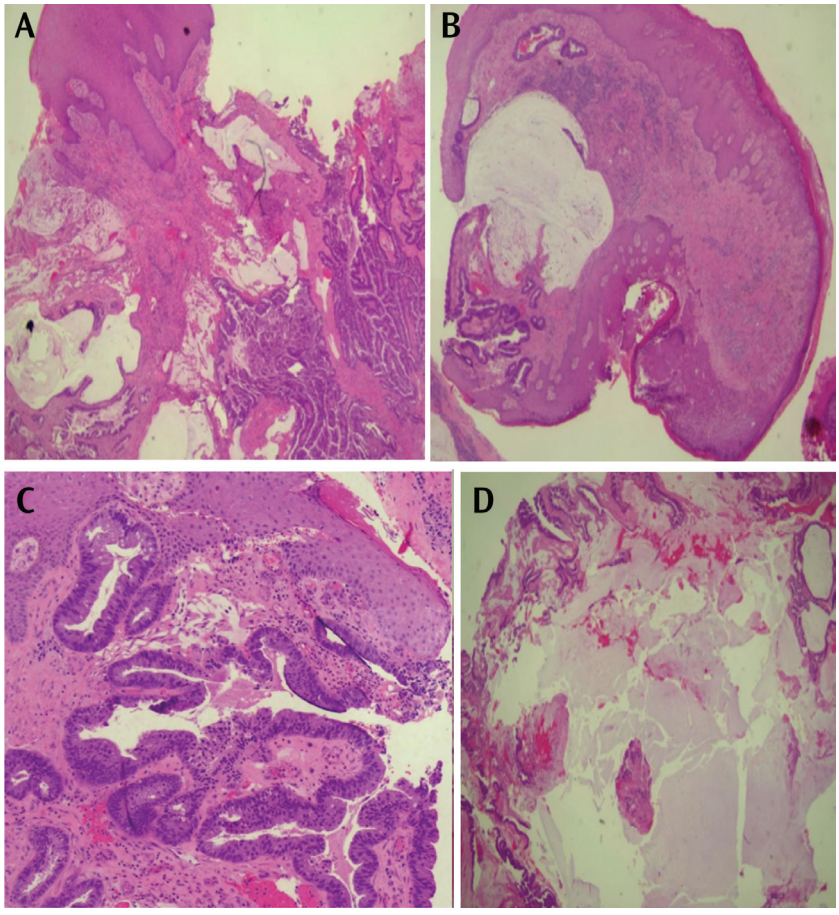

Figure 4. Carcinomatous glands infiltrating beneath inflamed squamous epithelium (A) and associated with mucinous production (B). (C) Malignant glands displaying moderate nuclear pleomorphism and loss of nuclear polarity. (D) Malignant glands accompanied by extracellular mucin pools.

Ethics

Informed Consent: Written informed consent was obtained from the patient's relatives for their anonymised information to be published in this article.

Acknowledgments

The authors would like to thank the Director General of Ministry of Health Malaysia for allowing to publish this case as a case report. Special thanks to those who were actively involved in managing this patient.

Footnotes

Authorship Contributions

Surgical and Medical Practices: K.H.H., A.B.K., Concept: F.H., Design: F.H., Analysis or Interpretation: A.B.K., Literature Search: F.M.A.H., Writing: F.M.A.H.

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

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

REFERENCES

1. Tekbaş A, Mothes H, Settmacher U, Schuele S. Non-mucinous adenocarcinomas and squamous cell carcinomas of the anal region masquerading as abscess or fistula: a retrospective analysis and systematic review of literature. J Cancer Res Clin Oncol. 2022; 148: 1509-22.
2. Chandramohan K, Mathew AP, Muralee M, Anila KR, Ramachandran K, Ahamed I. Squamous cell carcinoma arising from long-standing perianal fistula. Int Wound J. 2010; 7: 515-8.
3. Yamada K, Miyakura Y, Koinuma K, Horie H, Lefor AT, Yasuda Y, et al. Primary and secondary adenocarcinomas associated with anal fistulae. Surg Today. 2014; 44: 888-96.
4. Rosser C. The relation of fistula-in-ano to cancer of the anal canal. Trans Am Proc Soc. 1934; 35: 65-71.
5. Ghosh NK, Kumar A. Chronic fistula in ano associated with adenocarcinoma: a case report with a review of the literature. Ann Coloproctol. 2024; 40(Suppl 1): S1-5.

Gazi Med J 2026;37(4):576-582

DOI: http://dx.doi.org/10.12996/gmj.2026.4680

A Case of Acromegaly Diagnosed After Pituitary Apoplexy Presenting as Polycystic Ovary Syndrome

Polikistik Over Sendromu gibi Prezente Olan ve Hipofiz Apopleksisi Sonrasında Tanı Konulan Bir Akromegali Olgusu

Seyit Murat Bayram, Özgür Özçelik, Hüseyin Demirci, Revah Çankaya, Erman Çakal

Department of Endocrinology Clinic, University of Health Sciences Türkiye, Ankara Etlik City Hospital, Ankara, Türkiye

ABSTRACT

Pituitary apoplexy (PA) is an uncommon neuroendocrine emergency resulting from hemorrhage or infarction of a pituitary adenoma. Dopaminergic agents are well-established precipitants; however, the role of synthetic progestins remains insufficiently characterized. Acromegaly may present with clinical features that overlap considerably with those of polycystic ovary syndrome (PCOS), thereby complicating and delaying its diagnosis. We report a 26-year-old woman with a two-year history of secondary amenorrhea and hirsutism who had previously been evaluated by gynecology and diagnosed with PCOS on the basis of oligomenorrhea, clinical hyperandrogenism, and polycystic ovarian morphology on ultrasonography. Markedly elevated prolactin (127 µg/L) and a low free T4 level were also documented; however, the possibility of an underlying pituitary pathology was not investigated at that time. Dydrogesterone was commenced, and twelve days later the patient developed thunderclap headache and acute complete visual loss. Emergency evaluation demonstrated markedly elevated insulin-like growth factor-1 (984 ng/mL) and growth hormone (20 µg/L). Pituitary magnetic resonance imaging revealed a 50×25 mm macroadenoma with signal characteristics consistent with acute hemorrhage and infarction. Emergency transsphenoidal surgery was performed, and histopathology confirmed a sparsely granulated somatotroph pituitary neuroendocrine tumor of the Pit-1 lineage (Ki-67: 8%). The postoperative course was complicated by panhypopituitarism and permanent bilateral visual loss. This case underscores the importance of excluding sellar pathology before

ÖZ

Hipofiz apopleksisi (HA), hipofiz adenomunun kanaması veya enfarktüsü sonucunda gelişen, nadir görülen bir nöroendokrin acil durumdur. Dopaminerjik ajanlar iyi bilinen tetikleyici faktörler arasında yer almakla birlikte, sentetik progestinlerin bu süreçteki rolü henüz yeterince aydınlatılamamıştır. Akromegali, polikistik over sendromu (PKOS) ile belirgin klinik örtüşme gösteren bulgularla prezente olabilmekte; bu durum tanıyı güçleştirmekte ve geciktirmektedir. İki yıldır süren sekonder amenore ve hirsutizm şikayetiyle başvuran 26 yaşındaki kadın hasta, daha önce jinekoloji tarafından değerlendirilmiş; oligomenore, klinik hiperandrojenizm ve ultrasonografide polikistik over morfolojisi temelinde PKOS tanısı almıştı. Belirgin düzeyde yüksek prolaktin (127 µg/L) ve düşük serbest T4 değeri saptanmış olmakla birlikte, altta yatan hipofizer patoloji olasılığı o dönemde araştırılmamıştı. Didrogesteron başlanmasının ardından on ikinci günde hasta gök gürültüsü baş ağrısı ve akut tam görme kaybı ile başvurdu. Acil değerlendirmede belirgin derecede yüksek insülin benzeri büyüme faktörü-1 (984 ng/mL) ve büyüme hormonu (20 µg/L) düzeyleri saptandı. Hipofiz manyetik rezonans görüntülemesinde akut kanama ve enfarktüsle uyumlu sinyal özelliklerine sahip 50×25 mm'lik bir makroadenom izlendi. Acil transsfenoidal cerrahi uygulandı; histopatoloji, Pit-1 soyundan seyrek granüllü somatotrop hipofiz nöroendokrin tümörü tanısını doğruladı (Ki-67: %8). Postoperatif dönem panhipopituitarizm ve kalıcı bilateral görme kaybı ile komplike oldu. Bu olgu; menstrüel düzensizliği ve hafif endokrin bozuklukları olan hastalarda hormonal tedaviye başlamadan önce sellar patolojinin dışlanmasının önemini vurgulamaktadır. Ayrıca

Cite this article as: Bayram SM, Özçelik Ö, Demirci H, Çankaya R, Çakal E. A case of acromegaly diagnosed after pituitary apoplexy presenting as polycystic ovary syndrome. Gazi Med J. 2026;37(4):576-582

Address for Correspondence/Yazışma Adresi: Seyit Murat Bayram, Department of Endocrinology Clinic, University of Health Sciences Türkiye, Ankara Etlik City Hospital, Ankara, Türkiye
E-mail / E-posta: drmurat.endokrin@yandex.com
ORCID ID: orcid.org/0000-0003-1459-5765

Received/Geliş Tarihi: 02.04.2026
Accepted/Kabul Tarihi: 08.09.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

initiating hormonal therapy in patients with menstrual irregularities and subtle endocrine disturbances, particularly when features suggestive of acromegaly are present. It also highlights the need for broader awareness of acromegaly's clinical manifestations among non-endocrinology specialists.

Keywords: Pituitary apoplexy, acromegaly, dydrogesterone, PCOS

ÖZ

endokrinoloji dışı uzmanlık alanlarında akromegalinin klinik bulgularına yönelik farkındalığın artırılması gerektiğine dikkat çekmektedir.

Anahtar Sözcükler: Hipofiz apopleksisi, akromegali, didrogesteron, PCOS

INTRODUCTION

Pituitary apoplexy (PA) is a potentially life-threatening clinical syndrome caused by acute hemorrhage or infarction of a pituitary adenoma. It presents as a neuroendocrine emergency characterized by the sudden onset of severe headache, visual disturbances, ophthalmoplegia, and altered consciousness, frequently accompanied by acute hypopituitarism (1). Although the term "PA" classically refers to this dramatic clinical presentation, subclinical or silent apoplexy—detected incidentally on imaging or histopathological examination—is in fact considerably more prevalent (2). The estimated incidence of clinically overt PA is approximately 0.17 episodes per 100,000 individuals per year, occurring in 2-12% of patients with known pituitary adenomas (3). The pathophysiology of PA, while incompletely understood, is thought to involve a combination of factors: A mismatch between the elevated metabolic demands of a rapidly enlarging tumor and its blood supply, intrinsic vascular abnormalities within the adenoma, and compression of the superior hypophyseal vessels against the diaphragma sellae (4). Overexpression of vascular endothelial growth factor in certain adenomas has been identified as a contributing mechanism, potentially leading to the formation of fragile, fenestrated vessels prone to hemorrhage (5). A precipitating factor is identified in up to 30-40% of cases; recognized triggers include arterial hypertension, major surgery, dynamic pituitary function testing, coagulopathy, head trauma, and various pharmacological agents, of which dopamine agonists are the best-documented (6).

The association between sex steroid hormones and PA is less well established but has been reported. Two main pathophysiological mechanisms have been proposed. First, estrogens are known to induce pituitary hyperemia and increase glandular vascularity, potentially imposing hemodynamic stress on the fragile vasculature of an adenoma (5). Second, both estrogens and progestins increase the risk of thrombotic events by promoting a prothrombotic coagulation state, which may predispose to ischemic infarction within tumor tissue (7). While most reported cases have involved combined oral contraceptives, the independent contribution of synthetic progestins administered in isolation has not been systematically studied. Dydrogesterone is a retroprogestogen with a molecular structure closely resembling that of endogenous progesterone; it is widely used for the management of menstrual disorders, endometriosis, and luteal phase support owing to its favorable tolerability profile. Its potential to precipitate PA in the context of an undiagnosed pituitary macroadenoma remains inadequately documented, making individual case reports especially valuable for delineating this uncommon risk.

Acromegaly is a chronic disorder resulting from sustained excess secretion of growth hormone (GH), responsible for the characteristic peripheral manifestations, including progressive acral enlargement, coarsening of facial features, soft tissue hypertrophy, and a broad range of systemic complications. In more than 95% of cases, the source is a GH-secreting pituitary neuroendocrine tumor (PitNET). Because the onset is insidious and symptom progression is gradual, the diagnosis is frequently delayed by a decade or more (8). Consequently, approximately 75% of somatotroph adenomas are already at the macroadenoma stage (>1 cm) at the time of diagnosis, conferring greater potential for mass-related complications, including apoplexy (9,10). The co-occurrence of acromegaly and PA is uncommon; when it does occur, spontaneous biochemical remission of acromegaly may follow, but this is the exception rather than the rule, and most GH-secreting adenomas remain hormonally active after an apoplectic event (11,12). Critically, PA may occasionally represent the initial clinical presentation of a previously undiagnosed somatotroph adenoma.

We present a case of a young woman whose undiagnosed acromegaly, initially misattributed to PCOS by her gynecologists, was dramatically unmasked by the onset of PA shortly after dydrogesterone was initiated for secondary amenorrhea. This case illustrates the potential of synthetic progestins—and dydrogesterone in particular—to precipitate PA and emphasizes the critical importance of excluding sellar pathology prior to commencing hormonal therapy in patients with unexplained menstrual disturbances and hormonal abnormalities.

CASE PRESENTATION

A 26-year-old woman presented to the gynecology outpatient clinic in December 2024 with a two-year history of secondary amenorrhea and progressive hirsutism. She was a mother of three with no significant past medical history and no regular medications. Her body mass index was 35 kg/m^2, consistent with class II obesity. The preoperative gynecology notes documented obesity and non-specific constitutional complaints, without reference to acromegalic features. At the time of our endocrinology consultation—obtained following the apoplectic event—directed history taking revealed that the patient had noticed an increase in ring size and an increase in shoe size over the preceding two years, and physical examination demonstrated mild acral enlargement of the hands and feet. These findings, in retrospect, were consistent with chronic GH excess, though they had not been elicited or recognized at the time of the initial gynaecological assessment.

Laboratory investigations revealed a free thyroxine (T4) of 0.75 ng/dL (reference: 0.8-1.8 ng/dL), thyroid-stimulating hormone

(TSH) of 0.02 mIU/L (reference: 0.4-4.0 mIU/L), and a markedly elevated prolactin of 127 µg/L (reference: 3-27 µg/L) (Table 1). Transabdominal pelvic ultrasonography demonstrated bilateral polycystic ovarian morphology with a thin, regular endometrium. On the basis of secondary amenorrhea, clinical hyperandrogenism (hirsutism), and polycystic ovarian morphology on ultrasound—satisfying two of the three 2003 Rotterdam consensus criteria for PCOS (oligo-/anovulation and polycystic ovaries)—a clinical diagnosis of PCOS was established by the gynecology team (13). It is important to note that this diagnosis was made outside our endocrinology department; the patient was not referred to endocrinology at this juncture. The substantially elevated prolactin level, which raises the differential of a pituitary pathology such as a prolactinoma or a GH-secreting adenoma with co-secretion of prolactin, was not pursued further before initiating hormonal therapy. Dydrogesterone was commenced.

Five days after commencing dydrogesterone, the patient developed a headache and was advised to continue treatment. Twelve days after initiation, she returned with a markedly worsening headache and acute-onset visual loss, prompting an urgent referral to neurosurgery. Ophthalmological assessment at this point documented visual acuity of 0.25 in the right eye and absence of light perception (P-) in the left eye, with a relative afferent pupillary defect on the left. Fundoscopic examination showed normal optic discs bilaterally, and no acute ocular pathology was identified; retrobulbar neuritis and PA were raised as differential diagnoses by the ophthalmology team. Repeat laboratory investigations showed TSH 1.32 mIU/L, free T4 0.78 ng/dL, and prolactin 130 µg/L. Additional testing revealed a markedly elevated IGF-1 of 984 ng/mL (reference: 100-300 ng/mL) and GH of 20 µg/L (reference: <5 µg/L). Adrenocorticotropic hormone (ACTH) was 19 ng/L, and morning cortisol was critically low at 2.49 µg/dL (reference: 5-25 µg/dL), indicating adrenal insufficiency secondary to pituitary compromise.

Preoperative pituitary magnetic resonance imaging (MRI) demonstrated a 50×25 mm lesion occupying the pituitary fossa and extending into the sellar, suprasellar, and sphenoid sinus regions. The mass was in close proximity to both internal carotid arteries and the Circle of Willis and exerted significant compression on the optic chiasm, with superior extension to the level of the third ventricle. The lesion demonstrated heterogeneous signal characteristics on both T1- and T2-weighted sequences. On T2-weighted imaging, the presence of hypointense areas within the mass is consistent with the paramagnetic effects of deoxyhemoglobin or intracellular methemoglobin, indicating subacute hemorrhagic transformation. This pattern supports the interpretation of a hemorrhagic apoplectic event of recent onset, as hyperacute hemorrhage tends to appear isointense or hyperintense on T2 sequences, whereas the T2 hypointensity observed here is more characteristic of an evolving hemorrhage in the early subacute phase. The overall signal heterogeneity is consistent with mixed hemorrhagic and ischemic infarction within the adenoma (Figure 1).

Given the acute onset of complete visual loss and severe headache—both established indications for emergency surgical intervention in PA—the patient underwent emergency transnasal transsphenoidal pituitary surgery within hours of presentation. The decision to proceed with urgent decompression was guided by the presence of progressive neuro-ophthalmological compromise, in accordance with current management guidelines for PA (14). Published evidence further supports early surgery in patients with visual acuity impairment or visual field defects, as prompt decompression of the optic chiasm is associated with meaningful visual recovery (15).

Postoperatively, the patient was monitored in the intensive care unit. Day 1 laboratory results showed improvement in GH to 0.5 µg/L, prolactin below 2 µg/L, and IGF-1 at 554 ng/mL. However, hypernatremia developed (serum sodium 154 mmol/L, urine specific gravity 1.004), consistent with central antidiuretic hormone

Table 1. Preoperative and postoperative laboratory findings.

Parameter (reference range)	Pre-op (gynecology visit)	Pre-op (day before surgery)	Post-op day 1	Post-op day 3	Post-op week 1	Post-op month 1	Post-op month 6
Sodium (135-145 mmol/L)	—	139	154	145	143	143	—
Potassium (3.5-5.1 mmol/L)	—	4.2	3.6	3.2	—	—	—
Urine specific gravity (1.005-1.030)	—	—	1.004	1.020	—	1.025	1.020
TSH (0.4-4.0 mIU/L)	0.02	1.32	0.06	—	<0.005	<0.005	<0.005
Free T4 (0.8-1.8 ng/dL)	0.75	0.78	0.96	—	0.5	0.5	0.8
Cortisol (5-25 µg/dL)	—	2.49	4	—	0.8	0.2	0.5
ACTH (10-60 ng/L)	—	19	<1.5	—	<1.5	<1.5	<1.5
GH (<5 µg/L)	—	20	0.5	—	0.9	1.02	0.5
IGF-1 (100-300 ng/mL)	—	984	554	—	279	300	248
PRL (3-27 µg/L)	127	130	<2	—	<2	<2	<2
LH (2-12 IU/L)	—	<0.3	<0.3	—	<0.3	<0.3	<0.3
FSH (3-10 IU/L)	—	1	<0.3	—	<0.3	<0.3	<0.3
E2 (30-400 ng/L)	—	<5	16	—	<5	—	—

ACTH: Adrenocorticotropic hormone, E2: Estradiol, FSH: Follicle-stimulating hormone, GH: Growth hormone, IGF-1: Insulin-like growth factor-1, LH, Luteinizing hormone, Post-op: Postoperative, Pre-op: Preoperative, PRL: Prolactin, T4: Thyroxine, TSH: Thyroid-stimulating hormone. Note: Empty cells indicate that the parameter was not measured or that data were not available for that time point.

(ADH) deficiency. Anterior pituitary function remained severely impaired, with undetectable gonadotropins, TSH of 0.06 mIU/L, and morning cortisol of 4 µg/dL. Following endocrinology consultation, a diagnosis of panhypopituitarism with central ADH deficiency was established. Replacement therapy was initiated with levothyroxine, hydrocortisone, and desmopressin. By postoperative day 3, serum sodium had normalized to 145 mmol/L and urine specific gravity had increased to 1.020, indicating a therapeutic response.

Postoperative MRI (day 1) demonstrated a 31×22×18 mm postoperative cavity with peripheral enhancement, containing air and hemorrhagic material, and continuing to exert a mass effect on the optic chiasm. A 5×3 mm nodular area of enhancement adjacent to the left frontal base was observed, raising the possibility of residual tumor (Figure 2). Ophthalmological reassessment in the postoperative period documented bilateral fixed dilated pupils with no perception of light bilaterally (P-/P-) and bilateral optic nerve atrophy on fundoscopy, indicating irreversible ischemic injury to the anterior visual pathway.

One week post-surgery, TSH remained suppressed (<0.005 mIU/L), free T4 was 0.5 ng/dL, ACTH was below 1.5 ng/L, morning cortisol 0.8 µg/dL, and IGF-1 had decreased to 279 ng/mL. The patient was discharged on postoperative day 13 on desmopressin acetate 60 mcg twice daily, hydrocortisone 10 mg in the morning and 5 mg in the afternoon, and levothyroxine 25 mcg once daily.

Pathological analysis of the resected specimen identified a PitNET with positive immunostaining for synaptophysin, Pit-1, estrogen

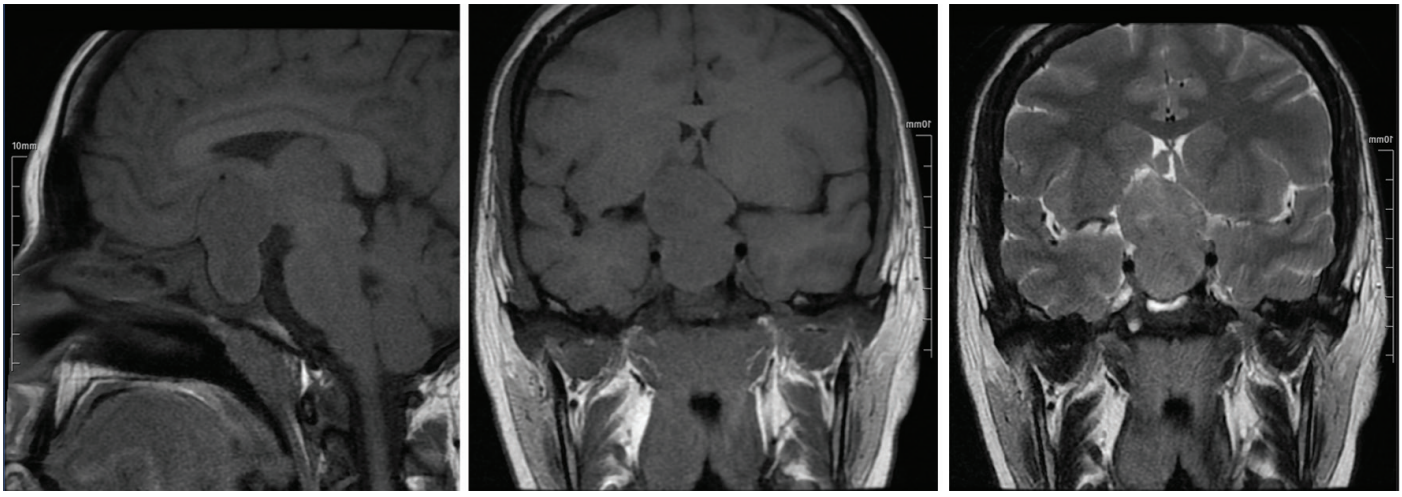

Figure 1. Preoperative pituitary MRI. From left to right: sagittal T1-weighted, coronal T1-weighted, and coronal T2-weighted images demonstrate a large macroadenoma (50×25 mm) filling the sella and suprasellar region. The lesion extends into the sphenoid sinus, compresses the optic chiasm, and abuts both internal carotid arteries. Heterogeneous signal intensity, including hypointense areas on T2-weighted imaging, is consistent with hemorrhagic transformation in the early subacute phase of apoplexy.

MRI: Magnetic resonance imaging.

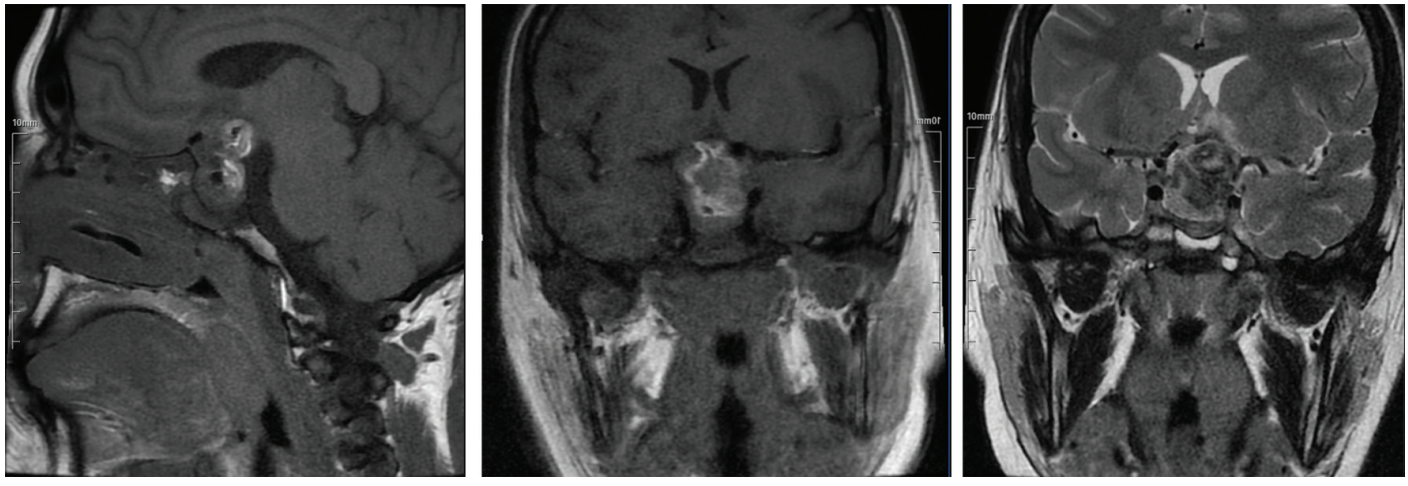

Figure 2. Postoperative day 1 pituitary MRI. From left to right: sagittal T1-weighted, coronal T1-weighted, and coronal T2-weighted images show a 31×22×18 mm postoperative cavity with peripheral enhancement, containing air and hemorrhagic material. A 5×3 mm nodular area of enhancement adjacent to the left frontal base raises the possibility of residual tumor. The optic chiasm remains compressed.

MRI: Magnetic resonance imaging.

receptor (ER), and GH. No staining for other pituitary hormones was detected. Cytokeratin 8-18 demonstrated a dot-like cytoplasmic pattern, and reticulin staining revealed disruption of the normal acinar architecture. The Ki-67 proliferation index was approximately 8%. The final pathological diagnosis was a sparsely granulated somatotroph PitNET of the Pit-1 lineage.

One month post-surgery, the patient remained on full pituitary hormone replacement with stable biochemistry. At four months, follow-up MRI showed the postoperative cavity had decreased in size to 20×14×13 mm and was partly occupied by fat packing material with peripheral enhancement. A 3×3 mm nodular area of enhancement adjacent to the right frontal base persisted, representing a possible residual tumor focus. The optic chiasm remained morphologically deformed (Figure 3).

At six months, anterior pituitary function remained globally deficient. The patient continued on her existing replacement regimen. GH was 0.5 µg/L, IGF-1 was 248 ng/mL, and all other hormonal axes remained suppressed (Table 1). She continues under combined outpatient follow-up by neurosurgery and endocrinology.

Written informed consent was obtained from the patient for publication of this case report and the accompanying clinical images.

DISCUSSION

This case illustrates how the insidious clinical presentation of acromegaly can result in diagnostic delays and misattribution to more common conditions such as PCOS, with potentially catastrophic consequences. The patient's presenting features of obesity, secondary amenorrhea, and hirsutism are well-recognized overlapping manifestations of both PCOS and acromegaly, and the clinical overlap between the two conditions has been described in the literature (16,17). A retrospective review of the preoperative gynecology documentation revealed subtle acromegalic features, including a reported increase in ring size and increase in shoe size over the preceding two years and mild indeterminate enlargement of the hands and feet on physical examination. Although non-specific in isolation, such findings—particularly in the context of markedly elevated prolactin and an isolated low free T4 with suppressed TSH, suggesting central rather than primary hypothyroidism—warrant systematic exclusion of pituitary pathology before initiating hormonal therapy. Non-endocrinology clinicians, including gynecologists and primary care physicians, should maintain awareness of the somatic features of acromegaly, as early recognition outside specialist settings is essential to reduce the well-documented diagnostic delay that characterizes this condition.

The markedly elevated prolactin level (127 µg/L) documented at initial presentation warrants specific discussion. When prolactin is substantially elevated in a patient with menstrual irregularities, the differential diagnosis should include prolactinoma, compression of the pituitary stalk by a non-functioning mass (the "stalk effect"), and co-secretion of prolactin by a GH-secreting adenoma. GH and prolactin co-secretion occurs in approximately 30-40% of acromegaly cases, either through a mammosomatotroph adenoma or a mixed somatotroph-lactotroph tumor, both of which fall within the Pit-1 transcription factor lineage (18). In the present case, the immunohistochemical profile of the resected tumor—strongly positive for GH, Pit-1, and ER, without prolactin staining—suggests that hyperprolactinemia was most likely the result of a stalk effect due to the large macroadenoma rather than true co-secretion. Irrespective of mechanism, a prolactin level of this magnitude in a young woman with menstrual disturbance should prompt pituitary imaging prior to the initiation of any hormonal treatment. The failure to investigate further before prescribing dydrogesterone was a critical missed opportunity (19).

The initial diagnosis of PCOS deserves careful consideration in this context. PCOS is the most common endocrinopathy in women of reproductive age, and its diagnosis is based on the 2003 Rotterdam consensus criteria, which require the presence of at least two of the following three features: oligo- and/or anovulation; clinical and/or biochemical signs of hyperandrogenism; and polycystic ovarian morphology on ultrasonography—after exclusion of other relevant endocrinopathies such as congenital adrenal hyperplasia, Cushing's syndrome, androgen-secreting tumors, hyperprolactinemia, and

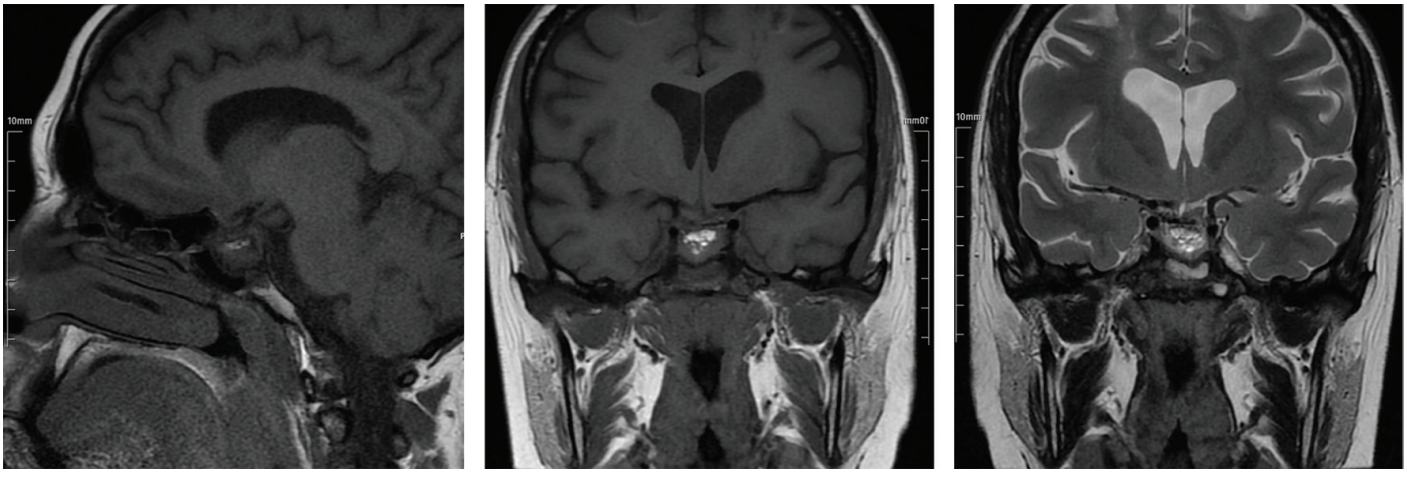

Figure 3. Postoperative month 4 pituitary MRI. From left to right: sagittal T1-weighted, coronal T1-weighted, and coronal T2-weighted images demonstrate a 20×14×13 mm postoperative cavity containing fat packing material with peripheral enhancement. A 3×3 mm nodular area of enhancement persists adjacent to the right frontal base, consistent with possible residual tumor. The optic chiasm remains morphologically deformed.

MRI: Magnetic resonance imaging.

thyroid dysfunction (13). Polycystic ovarian morphology, as defined by the Rotterdam criteria, requires the presence of 12 or more follicles measuring 2-9 mm in diameter in at least one ovary and/or an ovarian volume exceeding 10 mL. In the present case, the gynecology team identified oligomenorrhea and polycystic ovarian morphology on transabdominal ultrasound, satisfying two Rotterdam criteria. However, the exclusion steps integral to a rigorous PCOS workup—particularly the evaluation of markedly elevated prolactin and the abnormal thyroid function pattern suggestive of central hypothyroidism—were not completed. This serves as a reminder that PCOS is, by definition, a diagnosis of exclusion and that the presence of elevated prolactin or atypical hormonal patterns mandates further investigation before the diagnosis is confirmed and treatment initiated.

The temporal relationship between the initiation of dydrogesterone and the onset of PA in this case raises the possibility that this progestogen may have acted as a precipitating factor, though this association is based on temporal evidence rather than established causality. The precise mechanism by which sex steroids might trigger PA has not been fully elucidated. The proposed pathophysiological pathways include estrogen-mediated pituitary hyperemia with resultant vascular congestion and a procoagulant effect of exogenous progestins promoting thrombotic infarction within the adenoma (5,7). The majority of published cases linking hormonal agents to PA have involved combined oral contraceptives containing both estrogen and progestin components. Kobayashi et al. (20) reported a case of ischemic PA in a 33-year-old woman following 18 months of combined oral contraceptive use, proposing that the estrogen-induced hypercoagulable state was the underlying mechanism. More broadly, case series and reviews have documented PA in the setting of various sex steroid exposures, including combined oral contraceptives, anabolic androgenic steroids, and GnRH agonists, each mediated by distinct but potentially overlapping mechanisms (4,20-22).

The specific contribution of dydrogesterone as a sole precipitant of PA has not been established in the literature, and no controlled data exist. This case should therefore be interpreted cautiously: While the temporal sequence is compelling and consistent with a drug-related precipitant, a causal relationship cannot be confirmed from a single case report. The possibility that the apoplectic event was spontaneous within a large and hemorrhage-prone adenoma and coincidentally occurred shortly after dydrogesterone was started cannot be excluded. Despite extensive review of the available literature, no prior published case definitively attributing PA to dydrogesterone as an isolated agent was identified. The absence of evidence of causation does not, however, diminish the clinical lesson: In any patient with suspected pituitary pathology, hormonal agents capable of altering pituitary blood flow or coagulation status should not be initiated without adequate exclusion of a sellar mass.

The decision to proceed with emergency transsphenoidal surgery was made on the basis of acute-onset complete visual loss and thunderclap headache—both of which are accepted indications for urgent surgical decompression in the current management guidelines for PA. Published evidence supports early surgery (within 7 days) in patients with visual acuity impairment or visual field defects, as prompt decompression of the optic chiasm is associated with meaningful visual recovery in a significant proportion of patients (14,15). Unfortunately, in the present case, the severity of the apoplectic insult resulted in irreversible bilateral optic nerve atrophy with permanent loss of visual function, emphasizing the necessity of early recognition and rapid intervention.

Following surgery, the patient developed panhypopituitarism with central ADH deficiency, requiring comprehensive hormonal replacement with hydrocortisone, levothyroxine, and desmopressin. These complications are well recognized sequelae of PA, particularly after large apoplectic events, and reflect the extent of ischemic damage to the anterior and posterior pituitary parenchyma. Endocrine function did not recover over the six-month follow-up period, consistent with published reports indicating that pituitary insufficiency after PA is often permanent when the apoplectic injury is extensive (11).

To our knowledge, this case is the first reported case of a somatotroph macroadenoma presenting as PA potentially precipitated by dydrogesterone administered for a presumed diagnosis of PCOS. This case adds to a modest but growing literature on sex steroid-associated PA and highlights several important principles: The clinical overlap between acromegaly and PCOS mandates systematic pituitary evaluation in atypical cases; markedly elevated prolactin should always be investigated before commencing hormonal therapy; and the diagnosis of PCOS, as a diagnosis of exclusion, requires the formal exclusion of pituitary pathology. Non-endocrinology practitioners should be familiar with the somatic features of acromegaly—including acral enlargement, coarsening of facial features, change in ring or shoe size, soft tissue hypertrophy, and macroglossia—as their recognition in primary or specialist gynecological care has the potential to prevent delayed diagnosis and its consequences.

This case illustrates how an undiagnosed somatotroph macroadenoma presenting with clinical features mimicking PCOS was dramatically revealed by PA following the initiation of dydrogesterone therapy. It underscores the diagnostic challenges inherent to acromegaly; the potential for synthetic progestins to precipitate apoplexy in the presence of a pre-existing adenoma—though causality in this individual case cannot be confirmed—and the critical need for comprehensive pituitary evaluation before commencing hormonal treatment in patients with menstrual disturbances and unexplained endocrine abnormalities. Clinicians across specialties should maintain a low threshold for suspecting pituitary pathology when features overlap with PCOS, particularly when prolactin is markedly elevated or when subtle somatic features of acromegaly are present. When clinical suspicion is raised, measurement of IGF-1 and dedicated pituitary imaging should be obtained prior to initiating any hormonal therapy. This case contributes to the limited literature on progestin-associated PA and reinforces the importance of careful risk stratification in clinical practice.

Ethics

Informed Consent: Written informed consent was obtained from the patient.

Acknowledgments

AI tools were used for language correction. The authors reviewed and edited all AI-assisted text and take full responsibility for the content of the manuscript.

Footnotes

Authorship Contributions

Surgical and Medical Practices: S.M.B., Ö.Ö., H.D., R.Ç., E.Ç., Concept: S.M.B., E.Ç., Design: S.M.B., H.D., Supervision: H.D., E.Ç., Patient Follow-up: S.M.B., Ö.Ö., R.Ç., Literature Search: S.M.B., R.Ç., Critical Review: H.D., E.C., Ö.Ö., R.Ç., Writing: S.M.B.

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

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

REFERENCES

1. Briet C, Salenave S, Bonneville JF, Laws ER, Chanson P. Pituitary apoplexy. Endocrine Reviews. 2015; 36: 622-45.
2. Bonicki W, Kasperlik-Załuska A, Koszewski W, Zgliczyński W, Wisławski J. Pituitary apoplexy: endocrine, surgical and oncological emergency. Incidence, clinical course and treatment with reference to 799 cases of pituitary adenomas. Acta Neurochir (Wien). 1993; 120: 118-22.
3. Raappana A, Koivukangas J, Ebeling T, Pirilä T. Incidence of pituitary adenomas in Northern Finland in 1992-2007. J Clin Endocrinol Metab. 2010; 95: 4268-75.
4. Iglesias P. Pituitary apoplexy: an updated review. J Clin Med. 2024; 13.
5. Cardoso ER, Peterson EW. Pituitary apoplexy: a review. Neurosurgery. 1984;14: 363-73.
6. Semple PL, Jane JA, Laws ER. Clinical relevance of precipitating factors in pituitary apoplexy. Neurosurgery. 2007; 61: 956-61; discussion 961-2.
7. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the women's health initiative randomized controlled trial. JAMA. 2002; 288: 321-33.
8. Zahr R, Fleseriu M. Updates in diagnosis and treatment of acromegaly. Eur Endocrinol. 2018; 14: 57-61.
9. Lavrentaki A, Paluzzi A, Wass JA, Karavitaki N. Epidemiology of acromegaly: review of population studies. Pituitary. 2017; 20: 4-9.
10. Fernandez A, Karavitaki N, Wass JA. Prevalence of pituitary adenomas: a community-based, cross-sectional study in Banbury (Oxfordshire, UK). Clin Endocrinol (Oxf). 2010; 72: 377-82.
11. Capatina C, Inder W, Karavitaki N, Wass JA. Management of endocrine disease: pituitary tumour apoplexy. Eur J Endocrinol. 2015; 172: R179-90.
12. Fraser LA, Lee D, Cooper P, Van Uum S. Remission of acromegaly after pituitary apoplexy: case report and review of literature. Endocr Pract. 2009; 15: 725-31.
13. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). Hum Reprod. 2004;19: 41-7.
14. Rajasekaran S, Vanderpump M, Baldeweg S, Drake W, Reddy N, Lanyon M, et al. UK guidelines for the management of pituitary apoplexy. Clin Endocrinol (Oxf). 2011; 74: 9-20.
15. Brown NJ, Patel S, Gendreau J, Abraham ME. The role of intervention timing and treatment modality in visual recovery following pituitary apoplexy: a systematic review and meta-analysis. J Neurooncol. 2024; 170: 469-82.
16. Joham AE, Tay CT, Laven J, Louwers YV, Azziz R. Approach to the patient: diagnostic challenges in the workup for polycystic ovary syndrome. J Clin Endocrinol Metab. 2025; 110: e2298-308.
17. Khiyami AM, Jokar TO, Abdallah HM, Gardner PA, Zenonos GA, Styer AK, et al. Polycystic ovary syndrome preceding the diagnosis of acromegaly: a retrospective study in 97 reproductive-aged women. Reprod Biol Endocrinol. 2023; 21: 14.
18. Araujo-Castro M, Biagetti B, Menéndez E, Novoa-Testa I, Cordido F, Rodríguez Berrocal V, et al. Predictors of therapeutic failure in GH and prolactin co-secreting pituitary adenomas. Endocr Connect. 2025; 14.
19. Petersenn S, Fleseriu M, Casanueva FF, Giustina A, Biermasz N, Biller BMK, et al. Diagnosis and management of prolactin-secreting pituitary adenomas: a Pituitary Society International Consensus Statement. Nat Rev Endocrinol. 2023; 19: 722-40.
20. Kobayashi J, Miyashita K, Tamanaha T, Kobayashi N, Iihara K, Nagatsuka K. Pituitary ischemic apoplexy in a young woman using oral contraceptives: a case report. J Stroke Cerebrovasc Dis. 2013; 22: e643-4.
21. Donegan D, Erickson D. Revisiting pituitary apoplexy. J Endocr Soc. 2022; 6: bvac113.
22. Raj R, Elshimy G, Jacob A, Arya PVA, Unnikrishnan DC, Correa R, et al. Pituitary apoplexy induced by gonadotropin-releasing hormone (GnRH) agonist administration for treatment of prostate cancer: a systematic review. J Cancer Res Clin Oncol. 2021; 147: 2337-47.

Gazi Med J 2026;37(4):583-595

DOI: http://dx.doi.org/10.12996/gmj.2026.4698

Antibiotic Resistance Risk and Ecosystem

Antibiyotik Rezistans Riski ve Ekosistem

Evren Doruk Engin[1], Ayşe Başak Engin[2], Atilla Engin[3]

[1]Department of Medical Microbiology, Dokuz Eylül University, Faculty of Medicine, İzmir, Türkiye
[2]Department of Toxicology, Gazi University, Faculty of Pharmacy, Ankara, Türkiye
[3]Department of General Surgery, Gazi University, Faculty of Medicine, Ankara, Türkiye

ABSTRACT

Antibiotics are among the most prescribed drugs in human medicine. However, antibiotic resistance is a continuing global human health problem. Ecosystems and clinical practices play equally important roles in the emergence, evolution, and spread of antibiotic resistance. High doses and inappropriate antibiotic use result in large amounts of antibiotics being released into the ecosystem, leading to the selection of large numbers of resistant bacteria and the spread of antibiotic resistance genes (ARGs). Accelerated selection of antibiotic-resistant strains and increasing numbers of antibiotic-resistant bacteria (ARB) result in increased infection-related patient mortality rates. Hosts of environmental resistomes are largely restricted to the core of the environmental microbiome. Because of the phylogenetic barriers to horizontal transfer of ARGs, gene transfer occurs more frequently within species than between species. Horizontal gene transfer (HGT) drives the evolution of untreatable "superbugs" by concentrating ARGs within individual cells. Decreasing the colonization capacity of exogenous ARB, increasing human resistance to colonization, reducing horizontal transmission of ARG, and reversing antibiotic resistance in ARB are the most important strategies to combat antibiotic resistance by preventing transmission of ARB and ARG to humans.

Keywords: Antibiotic resistance, ecosystem, horizontal gene transfer, mobile genetic element, marine organism

ÖZ

Antibiyotikler, tıbbi uygulamalarda en çok reçete edilen ilaç gruplarından biridir. Ancak antibiyotik direnci, halen küresel bir insan sağlığı sorunudur. Ekosistemler ve klinik uygulamalar, antibiyotik direncinin ortaya çıkmasında, evriminde ve yayılmasında eşit derecede önemli rol oynar. Yüksek dozlarda ve uygunsuz antibiyotik kullanımı, ekosisteme büyük miktarda antibiyotik salınımına ve çok miktarda dirençli bakterinin seçilimine ve antibiyotik direnç genlerinin (ARG) yayılmasına yol açar. Antibiyotik dirençli suşların hızla seçilimi ve artan sayıda antibiyotik dirençli bakteri (ARB), enfeksiyona bağlı hasta ölüm oranlarının artmasına neden olur. Çevresel direnç genlerinin konakçıları büyük ölçüde çevresel mikrobiyom ile sınırlıdır. ARG'lerin yatay transferine yönelik filogenetik engeller nedeniyle, gen transferi türler arasında değil, türler içinde daha sık gerçekleşir. Yatay gen transferi (HGT), ARG'lerini aynı hücrede bir araya toplayarak tedavi edilemeyen "süper mikropların" evrimini yönlendirir. Ekzojen ARB'lerin kolonizasyon kapasitesini azaltmak, insanlarda kolonizasyona karşı direnci artırmak, ARG'nin yatay bulaşmasını azaltmak ve ARB antibiyotik direncini tersine çevirmek, ARB ve ARG'nin insanlara bulaşmasının önlenmesinde ve antibiyotik direnciyle mücadelede en önemli stratejilerdir.

Anahtar Sözcükler: Antibiyotik direnci, ekosistem, yatay gen transferi, mobil genetik element, deniz organizmaları

Cite this article as: Engin ED, Engin AB, Engin A. Antibiotic resistance risk and ecosystem. Gazi Med J. 2026;37(4):583-595

Address for Correspondence/Yazışma Adresi: Ayşe Başak Engin, Department of Toxicology, Gazi University, Faculty of Pharmacy, Ankara, Türkiye
E-mail / E-posta: abengin@gmail.com
ORCID ID: orcid.org/0000-0003-1300-4644

Received/Geliş Tarihi: 27.04.2026
Accepted/Kabul Tarihi: 18.07.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

INTRODUCTION

Antibiotics are among the most prescribed drugs in human medicine. However, global estimates suggest that only 50% of global antibiotic use is appropriate (1). In the United States of America, the Centers for Disease Control and Prevention (CDC) analysis of data from the National Ambulatory Medical Care Survey and the National Hospital Ambulatory Medical Care Survey indicate that 30% of the 154 million antibiotics prescribed each year are considered unnecessary (2). When the antibiotics are widely used, the reservoirs of resistant genes are prevalent, thus these tend to spread through ecological niches due to complex interactions between people in different environments (3). Antibiotic resistance is a continuing global human health problem. Controversially, the healthcare facilities contribute significantly to the emergence, evolution, and spread of antibiotic resistance, however also the ecosystems have an equally important role in its spread (4). Disposal of household medical waste and unused antibiotics in landfills or in wastewater poses a significant environmental and public health hazard. A significant portion of these drugs consists of stable compounds that are difficult to remove in wastewater treatment plants (WTPs) and that accumulate in the environment. Antibiotics such as ciprofloxacin and trimethoprim, as well as antineoplastic drugs such as cyclophosphamide, have been isolated from these waters (5). The contamination of hospital and city sewage wastewater with antibiotics and cytostatic drugs poses a major environmental threat (6). Cytostatic drugs, which inhibit DNA replication and are genotoxic to actively dividing cells of algae and aquatic plants, affect all proteins involved in the regulation of growth rate, the cell cycle, and redox status. Prolonged exposure of organisms to these toxins lead to unforeseen changes at the molecular level for the entire ecosystem (7-9). When considering the reuse of treated wastewater and biological solids for agricultural purposes, mutagenic potentials such as alteration in the chromosomal aberration and micronucleus frequency are extremely important in terms of bacterial behavior as well as the extent of environmental pollution (10). While assessing the causal relationship between antibiotic resistance and clinical practice, numerous confounding factors exclusive to each drug and microorganism, like pathogen-pathogen and pathogen-host interactions, mutations, cross-resistance, resistance mechanisms, and cross-reservoir drivers, must be considered (2). Antimicrobial resistance is associated with an estimated 4.95 million deaths due to the bacterial infections and the substantial cost of this phenomenon to the global economy which may amount to US$ 100 trillion until 2050 if no action is taken (11). The aim of this review is to discuss the "vicious cycle of environmental resistance" in which high doses of antibiotics released into the ecosystem during antibiotic administration lead to the selection of numerous resistant bacteria and the spread of antibiotic resistance genes (ARGs) (Figure 1). This comprehensive effort selected reports from a set of systematic research studies—performed with robust, validated methods and considering the significant and interdependent nature of the antibiotic resistance, encompassing human, animal, and environmental dimensions—to outline environmental antimicrobial-resistant (AMR) bacterial pollution and to discuss mechanisms of resistance development and their environmental properties. In this context, 98 articles were extracted from published studies, which were obtained by searching with the combined keywords "antibiotic resistance, ecosystem, horizontal gene transfer (HGT), and mobile genetic element", and 66 of these were conducted in the last 10 years. Since combining antibiotics with nanoparticle (NP)-based carriers offers significant advantages in overcoming antibiotic resistance, which is a global problem today, a comprehensive clinical perspective is provided to encourage the integration of nanoantibiotics (nAbts) into clinical practice.

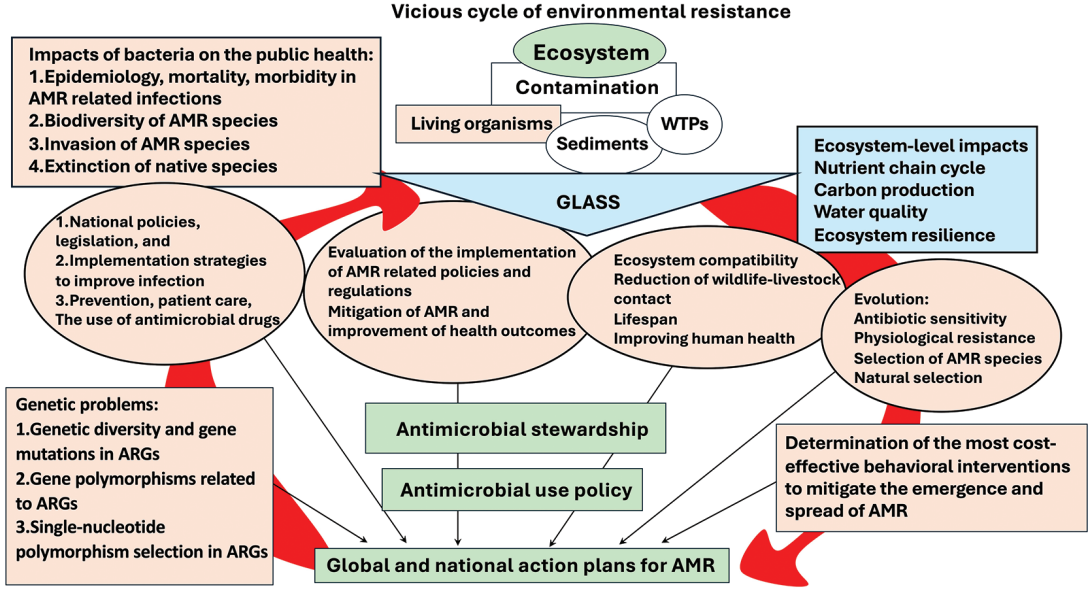

Figure 1. Vicious cycle of environmental resistance and the supporting factors of GLASS.

ARG: Antibiotic resistance genes, AMR: Antimicrobial resistance, GLASS: Global Antimicrobial resistance surveillance system - launched by World Health Organization in 2015, WTPs: Wastewater treatment plants.

Ecological Risk Assessment

Ecological risk assessment is based on the maximum measured concentration of pharmaceutical compounds in wastewater, expressed as a risk quotient (RQ), which is defined as the ratio of a point estimate of exposure [estimated environmental concentration (EEC) or predicted environmental concentration (PEC)] and a point estimate of effects (12). Compared with marine organisms, the maximum RQ values were 81,000 for ofloxacin and 440 for ciprofloxacin. These results indicate that antibiotics entering wastewater systems pose a high ecological risk due to improper wastewater treatment. This risk also means that waste pharmaceuticals can contaminate the food web and lead to antibiotic resistance in pathogenic bacteria (13). Twenty-two studies reporting drug concentrations in river and marine sediments have been published since 2000. A total of 62 drugs and their by-products were detected in marine precipitate at concentrations equivalent to 2,615,000 ng/g wet weight. Among these antibiotics, 26 were found to have the highest concentrations in the sediment (14). Fifty-nine papers published between 2002 and 2015 examined concentrations of 169 compounds in sludge from approximately 450 treatment systems, yielding data on 152 chemicals across 28 classes. Considering the environmental risk assessment of these compounds in soil, the most critical ones posing environmental risks in soil contaminated with sewage sludge were ciprofloxacin, ofloxacin and tetracycline (15). Fourteen studies reporting drug concentrations in shellfish measured tissue levels of 60 drugs and 7 by-products. Thirty-eight types of antibiotics constituted the most frequently reported contaminant class in these studies, followed by six types of antihypertensive agents. Ciprofloxacin and enrofloxacin were the most commonly reported antibiotics, each identified in at least four investigations. Only three reports reported the presence of erythromycin-based metabolites and its pharmaceutical transformation products (16,17). The effects of these pollutants were assessed in mussel tissues using biomarkers, such as activities of superoxide dismutase, catalase, glutathione reductase, and phase II glutathione-S-transferase (GST); lipid peroxidation levels (LPO), indicating oxidative stress; and acetylcholinesterase activity, associated with neurotoxicity. The results indicated potentially harmful impacts on species that may be sensitive to the continuous discharge of active pharmaceutical ingredients (APIs) into the ecosystem. Some compounds were found to significantly increase LPO levels in the liver, while others were found to increase GST levels in the gills. Indeed, APIs are commonly detected as environmental contaminants in surface waters due to incomplete and inappropriate wastewater treatment processes (18-20). Interestingly, however, some drugs were determined in marine organisms despite they could not be found in the sampled waters or sediments at concentrations below those expected to produce toxic effects (17).

Antibiotics and Their Ecological Effects

The use of antibiotics without a prescription or as a result of inappropriate prescribing leads to increased AMR, treatment failures, drug toxicity, and increased treatment costs (21). The World Health Organization (WHO) has identified AMR as one of the 10 most important threats to global health (22). A clear example of this occurred during the Coronavirus disease-2019 (COVID-19) pandemic in 2020. As an inevitable consequence of the increase in inappropriate antibiotic usage during the COVID-19 pandemic, AMR rates have increased significantly (23). In parallel, the heavy contamination of sewage with antibiotics has accelerated the selection of antibiotic-resistant strains within the bacterial population. This increasing number of antibiotic-resistant bacteria (ARB) results in increased infection-related mortality rates of patients (24). These suggest that antibiotics play a significant role in drug contamination of environmental waste and in their ecological impacts. Thus, erythromycin-base has been detected at concentrations of 3.4 ng/g in San Francisco Bay, United States, and 0.7-14 ng/g in the Pearl River, China (25). Furthermore, in waters sampled in Southern China, 27 antibiotics, primarily quinolones (6.36-463 ng/L) and aminoglycosides (94.9-458 ng/L), were identified as posing ecological risks. Higher concentrations of antibiotics are observed in river water during the dry season than during the wet season. Organic matter and ammonia nitrogen in the sediment strongly influence the distribution of these antibiotics. Nevertheless, the levels of antibiotics detected in this region are not considered significant in terms of ecological risk (26). However, in the analysis of sediment samples from the Liao River basin in China, 19 different antibiotics were identified, with tetracyclines and macrolides being the predominant antibiotics in terms of concentration. Quinolones were the most frequently detected antibiotic group, present in 57.1-100% of biological samples, with concentrations ranging from 286.6 to 1655.3 µg/kg. Among the quinolones, enrofloxacin had the highest bioaccumulation factor (BAF) that can be defined as the ratio between the concentration of the analyzed antibiotic in an organism exposed from all sources and the concentration of the same antibiotic in water, during steady state (27,28). Its pollution increases public concern about human health and the safety of aquatic ecosystems. Enrofloxacin in sediment is absorbed by lipid-rich fish. Therefore, enrofloxacin bioaccumulation in fish habitats is positively related to fish lipid content (29). Ciprofloxacin, which WHO classifies as a critically important antibiotic, was investigated in fresh surface water, groundwater, saltwater, treated municipal wastewater, untreated municipal sewage, treated hospital wastewater, and untreated hospital wastewater. Among the 260 articles reporting ciprofloxacin contamination in watersheds, 111 studies were from Europe, 105 from Asia, and 32 from North America. Of these, 5 studies were conducted in South America and 7 in Africa. To determine the "probability of exceeding the observed ecotoxicological effect concentration" such as Predicted No-Effect Concentration (PNEC) and "antibiotic resistance rate" (ABR) for antibiotics in wastewater "probabilistic hazard assessments" are conducted. The PNEC is the concentration of an antibiotic in any environment below which adverse effects will most likely not occur during long-term or short-term exposure. The risk of antibiotic resistance (PNEC-ABR) is assessed by comparing the PNEC with the EEC (or PEC) or with the minimum effective concentration for the development or maintenance of antibiotic resistance. Since, there is no universally accepted or regulatory algorithm for obtaining a reliable ABR-PNEC for bacteria, it is estimated by applying evaluation factors to bacterial minimum inhibitory concentrations (MICs) (12). Remarkably, across all geographic regions, 58% of municipal effluents exceeded the ABR-PNEC for ciprofloxacin. Concentrations reported in surface waters exceeded the ABR-PNEC by 25%. Concentrations, which exceeded the ABR-PNEC were found to be by 31% in Europe,

29% in North America, and 67% in Africa (30). The interaction of ciprofloxacin with activated sludge from sewage treatment plants influences its migration and the control of associated risks. Sewage sludge removes over 96% of the ciprofloxacin from water through absorption, thereby increasing the ciprofloxacin concentration in the sediment. Extracellular proteins interact with ciprofloxacin. Ciprofloxacin contamination results in a decrease in soluble protein level, a significant elevation in extracellular protein concentration, and the formation of new extracellular proteins (31). The use of sewage sludge in ecological agriculture and the consumption of its products by humans pose a danger to human health by allowing humans to ingest products contaminated with antibiotics and increasing the resistance of commensal bacteria to these antibiotics. As an example, erythromycin has been detected in 1) wastewater, drinking water, groundwater, estuarine and coastal systems; 2) sewage sludge, biosolids, and sediments; and 3) tissues of aquatic organisms. In a study in China, 72.4% of surface waters were found to be contaminated; in a study in the United States, 85.4% were found to be contaminated; and in a study in Spain, 41.9% were found to be contaminated. The PNEC-ABR for erythromycin was found to be 33.3% higher in Asia, 20% in Europe, and 17.8% in North America (32). Antibiotic uptake by marine organisms, which are crucial for human nutrition, is specific to the compound type, the organism species, and the body tissues. Oxytetracycline accumulates in the organs, while oxolinic acid (a quinolone family antibiotic that inhibits DNA gyrase) accumulates in the gills (33). Indeed, among mollusks species collected from the Bohai Sea of China, antibiotic uptake appeared to vary depending on the species (34). High concentrations of fluoroquinolones have been found in liver and muscle tissue in fish from the marine aquaculture areas of the Pearl River Delta in South China (35). WTPs constitute a large reservoir of antibiotics, ARGs, and mobile genetic elements (MGEs). Mechanical and biological wastewater treatment methods are insufficient for removing these micro-pollutants. Exposure of ARB to heavy metals in the soil exacerbates adverse environmental impacts and accelerates the spread of resistance mechanisms. Excessive heavy metals in the environment and their high pressure on microorganisms directly affect the biodiversity of microbial communities living in the ecosystem (36). HGT is a significant ecological risk factor, leading to the co-selection of ARGs and metal resistance genes (MRGs) due to their co-localization on the same MGEs. Resistance arising from genetic changes triggered by environmental heavy metals in bacteria entering the environment through sewage systems and WTPs is an ecological factor that significantly reduces the effectiveness of antibiotics against bacterial infections (37). Significant correlation between abundance of ARGs and MGEs in a total of 164 ARGs from 50 soil samples, revealed the importance of HGT for ARG dissemination (38). HGT and co-localization of the host's MGE and MRG resistomes are shared between animal and human fecal samples. Many ARGs are shared by animals and human feces, in addition to the MGEs-containing class 1 integron, which are identified on animal feces, are carried by human pathogens (39).

Ecosystem and Antimicrobial Resistance

The rapid development of ARB is a global health security emergency. Current attention is centered on the mechanisms of transmission of ARB across species and the identification of environmental reservoirs of antibiotic resistance in wastewater (40). To date, estimates disclose that 700,000 deaths occur each year due to the antimicrobial resistance; afterward, by 2050, 10 million deaths are expected every year (41). The genes selected are widely detected in wastewater samples, indicate various resistance mechanisms, and are as well located in different genetic elements, specifically in MGEs that substantially affect the spread of ARGs (42).

Resistance to antibiotics is acquired by bacteria via two routes: either vertical evolution (endogenous) or horizontal evolution (exogenous). Vertical evolution requires a spontaneous mutation in the bacterial genome that confers increased resistance to an antibiotic. Horizontal evolution occurs through the transfer of a resistance gene from a resistant bacterium to a susceptible one. The mechanisms of this route are conjugation, transduction, and transformation. During conjugation, the resistance plasmids containing ARGs are transferred between bacteria. Transformation is a kind of modification of the bacterial genome by the uptake and incorporation of exogenous DNA, while the transduction mechanism is the transfer of bacterial DNA via a viral vector (43). Wildlife contributes to the spread of antibiotic resistance across a variety of hosts and ecosystems at two levels. First, this transmission occurs locally, where animals living alongside humans can serve as reservoirs and vectors for the spread of antibiotic resistance. Second, migratory animals such as seagulls, fish and turtles contribute to the spread of antibiotic resistance across geographic regions and even between continents, posing a global health concern (4). Furthermore, the ARGs can be transferred between distantly related bacterial species and to the human pathogens and the bacteria that are capable of colonizing the human body (44). Ecological factors are regarded as potential hotspots for the spread of antimicrobial resistance. Antimicrobial contamination not only increases the resistance among environmental bacteria, but also directly impacts the virulence of antimicrobial-susceptible bacterial infections in patients by enhancing MICs of antibiotics and bacterial transmission (45). The proliferation of ARB results in alterations in the antibiotic susceptibility of the overall microbial population when antimicrobial levels are below the MIC. Consequently, resistant bacteria generally survive even in the availability of an antimicrobial level, which is sufficient to inhibit or kill microorganisms of the same species (46). Total enterococci and vancomycin-resistant enterococci are detected in 71% (34/48) and 4% (2/48) of reclaimed water samples, respectively, while *Enterococcus faecalis (E. faecalis*) is the major species. These enterococci in the reclaimed water present high antibiotic resistance at the point of water use as determined by using their MICs (47). When tomatoes were exposed to sulfamethoxazole and trimethoprim, they accumulated the antibiotics at 5.26 µg/kg and 3.40 µg/kg, respectively. Since the accumulated antibiotics, ARGs, and antibiotic-resistant pathogens (ARPs) are ingested via food, treatment of ARPs due to antibiotic resistance poses a significant challenge (48). Enrofloxacin, ciprofloxacin, ofloxacin, oxytetracycline, and erythromycin pose a high risk of selecting for antimicrobial resistance in marine culture water and/or sediment. Sulfonamide resistance genes *sul1* and *sul2* are the most common ones in water and sediment, whereas tetracycline resistance genes are widespread in the gills and intestines of fish. Species of *Proteobacteria, Actinomyces*, and *Cyanobacteria* are positively correlated with ARGs, suggesting that they may be potential hosts for

ARGs. Antibiotic resistance and human infection-related functions of the microbiome are significantly higher in fish than in marine culture (49). The development of antibiotic resistance in marine bacteria has been associated with antibiotic concentrations in wastewater discharges and aquaculture. Some recently identified genetic elements and resistance determinants for quinolones, tetracyclines, and β-lactamases have been found to be shared among aquatic bacteria, fish pathogens, and human pathogens. To demonstrate the ecological effects, 35 isolated bacterial strains were tested for resistance to the eight most frequently detected antibiotics. Resistance to the tested antibiotics was detected in 97.1% of the bacterial strains, of which 70.6% exhibited high-level resistance to the same antibiotics, particularly sulfonamides. Consequently, high bacterial resistance to antibiotics was significantly correlated with average antibiotic concentrations (50,51). ARB strains have been reported to be more prevalent in marine wildlife populations exposed to sewage contamination. Sharks sampled in the Florida Keys exhibited resistance to a broad range of antibiotics. Resistance to at least one antibiotic was detected in all fish species sampled from each of the six study sites. Multidrug resistance was also reported in most of the evaluated sites. Interspecific comparisons among redfish, *Sciaenops ocellatus,* and sharks from Louisiana offshore waters revealed a significantly higher prevalence of resistance depending on the age of the population (52). Evidence has emerged suggesting that ARB found in seabirds are of human origin. Several extended-spectrum β-lactamase-producing *Escherichia coli* (*E. coli*) isolated from birds clustered with human-derived isolates. Therefore, wild birds acquire *E. coli* with human-derived and human-resistant traits and thus, serve as an environmental reservoir with the potential to reinfect human populations (53).

The presence of ARGs in marine ecosystems indicates ecological changes caused by antibiotics. The presence of ARB phenotypes was investigated by direct antibiotic susceptibility testing in surface waters collected from 12 sources. According to the antibiotic susceptibility test results, resistance ranged from 17% to 100%. The antibiotic resistance profile against a total of 19 antibiotics was as follows: Amikacin (17%), gentamicin (17%), ciprofloxacin (33%), colistin (42%), linezolid (42%), tobramycin (42%), vancomycin (42%), ertapenem (67%), erythromycin (42%), meropenem (75%), rifampicin (75%), teicoplanin (75%), tetracycline (75%), trimethoprim (75%), fusidic acid (83%), aztreonam (92%), clindamycin (92%), penicillin (92%), and cefoxitin (100%). Multiple resistance, ranging from one to six antibiotic classes, was observed. It was found that antibacterial activity decreased sequentially starting from aminoglycosides to fluoroquinolones, glycopeptides, macrolides, tetracyclines and β-lactams (54). Although *E. coli* are inherently sensitive to practically all clinically significant antimicrobial agents, these bacteria have a considerable capacity to acquire resistance genes, usually through HGT. *E. coli of* animal origin often shows resistance to other, frequently older, antimicrobial agents, such as tetracyclines, phenicols, sulfonamides, trimethoprim, and fosfomycin. However, multi resistance plasmids and other MGEs, like transposons and gene cassettes in class 1 and class 2 integrons have significant affect in the dissemination of resistance genes among *E. coli* species (55). Evidence indicates that submarine discharges of treated effluents have negative impacts. The overall ARG content in the investigated microbiomes was dominated by genes encoding multidrug efflux pumps, including MexAB-OprM, AcrEF-TolC, and MdtEF-TolC. The others are the modified peptidoglycan precursors, transporter genes conferring resistance to tetracycline, macrolide, and phenicol antibiotics, and the bla operon conferring β-lactam resistance. Several pathogenic genera introduced by effluents, like *Acinetobacter, Arcobacter, Bacteroides, Escherichia-Shigella, Klebsiella, Pseudomonas*, and *Salmonella*, show efflux pump-driven multidrug resistance, however *Acinetobacter, Salmonella, Bacteroides* and *Pseudomonas* emerge as predominant carriers of non-efflux ARGs conferring resistance to most of nine antibiotic classes (56). The core marine resistome is primarily involved in glycopeptide and multidrug resistance. Hosts for this resistome are largely restricted to the core marine microbiome, and gene transfer occurs more frequently within rather than between species, because of the presence of the phylogenetic barriers against the horizontal transfer of ARGs (57). Amongst multiple genetically polluted environments, water environments are the most likely ones constituting huge ARG reservoirs into which clinical and terrestrial bacteria flow and in which variety of human commensal bacteria develop (58).

Horizontal gene transfer

Bacteria have developed complicated strategies to counter pharmaceutical agents, such as decreasing antibiotic influx; reducing the drug's binding affinity or preventing drugs from effectively inhibiting the target enzyme; reducing intracellular concentration to sub-inhibitory levels via multidrug efflux pumps; inactivating the drug via hydrolysis of the antibiotic's functional group or transferring a chemical moiety to the antibiotic. These resistance mechanisms against antibiotics may be intrinsic to the species of bacteria or acquired from others via HGT (59). HGT, accomplished by conjugation by plasmids, transduction by bacteriophages, and natural transformation by extracellular DNA, allows genetic material to pass between strains and species to circumvent the normal vertical inheritance (60). In addition to these mechanisms, OMVs, which are abundant DNA reservoirs in aquatic environments, play an important role in the exchange of ARGs between bacteria (58,61). Of the three canonical mechanisms of HGT, conjugation is thought to have the greatest influence on the dissemination of ARGs (62). In fact, HGT is an epidemic infection model. The model is designed as a combination of quantitative measures of HGT and vertical gene transfer (VGT). It was determined that HGT is affected by the inhibitory mechanism of an antibiotic and its targets include donor, recipient alone or together, while VGT contributes vastly to the formation of transconjugants and consequently ARG spreading (63). In the environment, horizontal DNA transfer between multiple bacterial species and genera occurs by transformation and transduction, but primarily by conjugation. Conjugation is liable for the spread of genes coding for antibiotic resistance (64). Many environmental factors like nanomaterials, variety of oxidants, as well as light influence the horizontal transfer of ARGs in water environments (65). Furthermore, biofilm formation in water environments is considered to be hot spot for HGT of ARGs (58). HGT drives the evolution of untreatable "superbugs" by facilitating the accumulation of ARGs within the same cell. Thereby, HGT allows microbes to rapidly gain new capabilities and adapt to the environment. High rates of HGT have been detected among marine bacterial species. The frequency

of gene transfer detected in the oceans today is between a thousand- and a hundred-million-fold higher than previous estimates of HGT (60,66). ARGs are spread through HGT of ARG-associated MGEs (67). In addition, MGEs promote the assortment of AMR bacteria (68). Extracellular vesicles (EVs) are mediators of HGT and the spread of antibiotic resistance. In bacteria, the mrdA/mrdB pathway—a critical protein complex—consists of PBP2 (encoded by mrdA) and RodA (encoded by mrdB). Transcriptomic analysis has shown that mrdA and mrdB are key regulators of EV biogenesis. The mrdA/mrdB pathway has provided insights into the molecular mechanisms by which sub-inhibitory dose antibiotic exposure promotes the spread of antimicrobial resistance (69). New Delhi metallo-β-lactamase (NDM-1) is one of the most potent and widespread carbapenem-hydrolyzing enzymes; it is commonly found in bacteria and is secreted into the extracellular environment within outer membrane vesicles (OMVs). OMVs containing NDM-1 protect carbapenem-susceptible bacterial strains against meropenem. The secretion of NDM-1-containing OMVs ensures the survival of carbapenem-susceptible bacteria around the infection site (70). Indeed, OMVs from *E. coli* carrying NDM-1 not only display carbapenemase activity but also can protect the populations of carbapenem-susceptible bacteria (71). To reduce the spread of resistant strains and protect global public health, stronger antimicrobial governance and policy reforms are needed, including novel therapeutic approaches, such as phage therapy and nAbts. Effective interventions across multiple areas, from diagnosis to antibiotic administration policies, are critical to control the threat of antimicrobial resistance (72). On the other hand, studies support the possibility that ARGs in hospitals can spread through HGT among bacteria found in the microbiomes of animals such as cockroaches, fleas, or rats, thus these organisms may be bioreactors for the horizontal transfer of ARGs among human pathogens (4). ARGs in plasmids can be transmitted across clinical and environmental boundaries. Network analysis revealed four β-lactam genes (*blaTEM-1, blaNDM-4, blaKPC-2*, and *blaSHV-1*) among 1029 plasmids and 45 clinical pathogens. This put forward that clinically alarming ARGs transfer expedite the propagation of antibiotic resistance in clinical pathogens (73). Packaged gene sets transferred via HGT allow bacteria to exchange multiple ARGs in a single transfer event (74). Horizontal transfer of CTX-M β-lactamase plasmid-borne *blaCTX-M* genes detected in *Klebsiella pneumoniae* (*K. pneumoniae*) isolates from hospital environment contributes to the spread of *K. pneumoniae* and creates an epidemiological hazard (75). In summary, HGT plays a significant role in the microbial ecology of antimicrobial resistance, as it allows the exchange of ARGs among taxonomically similar bacteria, and HGT rate depends on environmental conditions. In this context, multidrug-resistant soil bacteria have perfect nucleotide identity of genes from various human pathogens (76).

Clinical Perspective

During therapeutic antibiotic use, prolonged exposure of bacterial pathogens to high concentrations of antibiotics exerts significant selective pressure, leading to the development of more resistant bacteria. Although new antibiotics have shown a significant reduction in the persistence of bacterial infections, the development of resistance and biofilm formation generally minimize antibiotic effectiveness (77). The main disadvantages of conventional antimicrobial agents are their contribution to the development of multidrug resistance and their adverse effects. The biggest problem, however, is that newly synthesized antimicrobials do not have the

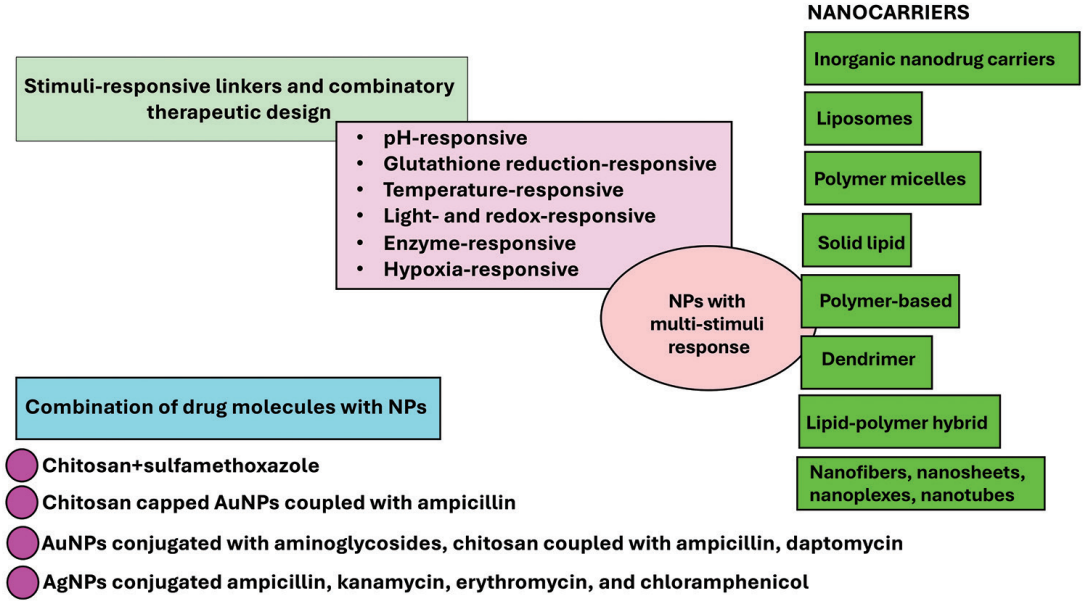

Figure 2. Types of nanoantibiotics (nAbts) that are currently used for drug delivery. The most significant advantage of nAbts, formulated to deliver drugs to infection sites using antibiotic-loaded nanocarriers, over traditional drug delivery systems is that nanoparticles (NPs) prevent the aggressive proliferation of resistant bacteria at the infection site by giving antibiotics novel functions through stimulus-triggered mechanisms.

Ag: Silver, Au: Gold, Sr: Stimuli responsive, nAbts: Nanoantibiotics, NP: Nanoparticle.

potential to provide a long-term solution (78). Combining antibiotics with NP-based carriers that offer better biofilm penetration, lower toxicity, and greater efficacy in combating bacterial resistance has become essential in overcoming antibiotic resistance, a global problem today. However, the clinical application of metal NP-based strategies against bacterial infections has been delayed (79) (Figure 2). In this context, the U.S. Food and Drug Administration (FDA)-approved- nAbts gold (Au) nanoclusters (NC), auranofin analogues are found to be effective against *Clostridium difficile* (*C. difficile*) infections, vancomycin-resistant *Enterococci*, methicillin-resistant *Staphylococcus aureus* (*S. aureus*) (MRSA) (auranofin has the capacity to eradicate the intracellular MRSA that is found inside infected macrophages), *Helicobacter pylori (H. pylori*) infections (synergistic with the amoxicillin via targeting both the reductive enzyme thioredoxin reductase and cell wall synthesis). Their anti-plasmodial activities likely result from direct inhibition *of Plasmodium falciparum* (*P. falciparum*) thioredoxin reductase. In addition, biofilm-associated bacteria can be treated with ultrasmall AuNPs (80). The daptomycin (Dap)-Au NCs hybrid structure demonstrates a stronger bactericidal effect toward MRSA which is a representative of multidrug-resistant bacteria, over the physically mixed Au NCs and daptomycin (Dap+Au NCs) (81).

It has been shown that the co-delivery of β-lactam antibiotics and β-lactamase inhibitors as nAbts, based on metal-carbenicillin framework-coated mesoporous silica NPs (MSN), is an effective method against MRSA. The constructed β-lactamase inhibitor-loaded nAbts are stable under physiological conditions. These nano formulations are capable of simultaneously releasing antibiotic molecules and the inhibitors at the site of bacterial infection. Thus, they achieve more effective elimination of ARB strains and biofilms. Ultimately, β-lactamase inhibitor-loaded nAbts exhibit greater penetration into biofilms, making them more effective at inhibiting MRSA (82). Folic acid tagged chitosan NPs are designed to function as "Trojan horses" to deliver vancomycin into bacterial cells. Since these NPs are biocompatible, biodegradable, semisynthetic polymers, they serve as valuable tools. Nanosized chitosan derivative loaded with vancomycin improves the transport of vancomycin across epithelial surfaces, thereby display antibacterial activity on vancomycin resistant *S. aureus* (83). Chitosan-capped AuNP coupled with ampicillin nAbt exhibited a two-fold increase in antimicrobial effectiveness compared with free ampicillin, while the antibiotic dosage was reduced to 50% of conventional usage (84). Ampicillin-functionalized AuNPs utterly demolish ampicillin-resistant bacteria (85). Similarly, the MIC of vancomycin-AuNPs against *S. aureus is* lower than that of free-form vancomycin. Therefore, the growth of intracellular *S. aureus* is successfully inhibited in vancomycin-AuNPs-engulfed macrophages (86). The combined antimicrobial effects of composite NPs and conventional antibiotics are more effective against bacteria resistant to conventional antibiotics. Although two different NPs are effective against the multidrug-resistant pathogens, each is difficult to functionalize with biomolecules and drugs. However, doxycycline-conjugated bimetallic NPs demonstrate substantially higher bactericidal activity against *Pseudomonas aeruginosa* (*P. aeruginosa*), *E. coli,* and *S. aureus*. This finding reveal that synergy of antibiotic with bimetallic NPs present a greatly efficient application (87). Compared to the ciprofloxacin, functionalized single-walled carbon nanotubes-ciprofloxacin show substantial increment in the antibacterial activity by 16-fold for *S. aureus* and *P. aeruginosa* and by 8-fold for *E. coli* (88). In the majority of bacterial infections, virulence factors account for the rapid proliferation of microorganisms within the host. Therefore, novel therapeutic nAbts are designed with diverse linkers to confer novel functionalities, such as pH-sensitivity, temperature-sensitivity, light-sensitivity, redox-sensitivity, and enzyme-sensitivity. Stimulatory mechanisms launched by NPs increase the selectivity of nAbts activation and the inhibition of bacteria within the biofilm (89,90) (Figure 2).

Considering that the antimicrobial mechanism of conventional antibiotics is generally simple, bacterial resistance emerges rapidly. Contrarily, majority of the traditional antibiotic resistance mechanisms are void for metal NP (mNP)-based strategies (91). Based on existing data, nAbts provide a novel type of attack against which the bacteria currently lack a defense mechanism. Despite the resembling antibacterial efficiency, nAbts display lower toxicity in comparison to free-drug administration (92). Additionally, the protein corona formation significantly influences the uptake, biotransformation and biocompatibility as well as the cytotoxicity of NPs (93-95) (Figure 3). Opsonization of therapeutic NPs by protein adsorption results in substantial removal by reticuloendothelial cells, decreasing the bioavailability of circulating NPs at the targeted application site. Significantly reducing the immunotoxicity of NPs makes the NP carriers beneficial for antibiotic delivery (96,97).

In fact, antibiotic susceptibility characteristics define the relationship between a microorganism's *in vitro* susceptibility to the drug and the *in vivo* exposure of the microorganism to the drug (98). Metal oxide NPs used in conjunction with antibiotics not only reduce the toxicity of antibiotics but also improve their antibacterial efficacy (99). Antibiotic-loaded nanocarriers are designed to deliver drugs to infection sites, and mNP carriers protect antibiotics against resistance mechanisms developed by target bacteria (100) (Figure 3). Within mNP carriers, drugs are kept safe from harmful chemical reactions until they arrive their target *in vivo*, thus those carriers preserve the efficacy of the drugs and from antibiotic resistance mechanisms (101).

NP-based antibacterial drug delivery systems deliver the drug directly to the target site, thereby side effects are reduced. By proper NP carrier-mediated drug release, blood levels of the antibiotic at the targeted infection site can be sustained at the required effective concentrations for a prolonged period. This leads to improved stability and reduced frequency of antibiotic administration. Compared with the free drug at the same concentration, the drug delivered via an NP carrier has been shown to have a much greater inhibitory effect on cellular growth and is accompanied by prolonged drug release (102) (Figure 4). nAbts are administered alone or in combination with antibiotics to increase the likelihood of biofilm eradication or shorten the duration of treatment. Since polysaccharides and extracellular DNA (eDNA) associated proteins in the extracellular polymeric substance (EPS) matrix participate in the biofilm structural integrity, deoxyribonucleases (DNases) disrupt biofilms (103). EPS matrix-degrading enzymes facilitate dispersion of bacteria within biofilms, leading to more effective bacteriolysis when combined with antimicrobial agents. Dispersin, glucanohydrolases, glycoside hydrolases, and DNases enhance delivery of antimicrobial

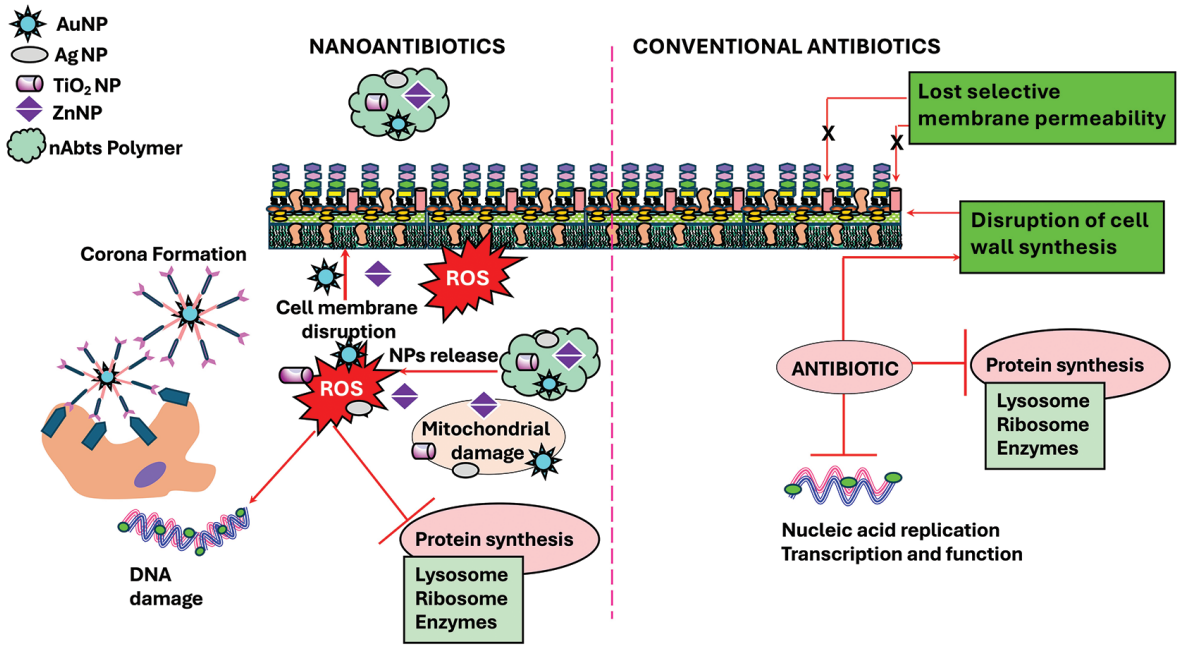

Figure 3. Conventional antibiotics disturb the bacterial membrane permeability and synthesis, as well as inhibit nucleic acid transcription and function. In contrast, nanoantibiotics (nAbts) act by impairing bacterial membranes via generating reactive oxygen species (ROS), leading to oxidative stress, disruption of transmembrane electron transport, and damage to mitochondria and DNA. The combination of antibiotics and metal oxide nanoparticles (mNPs) has been shown to reduce antibiotic toxicity while increasing their antibacterial efficacy.

Ag: Silver, Au: Gold, nAbts: Nanoantibiotics, NP: Nanoparticle, ROS: Reactive oxygen species, TiO_2: Titanium dioxide, Zn: Zinc.

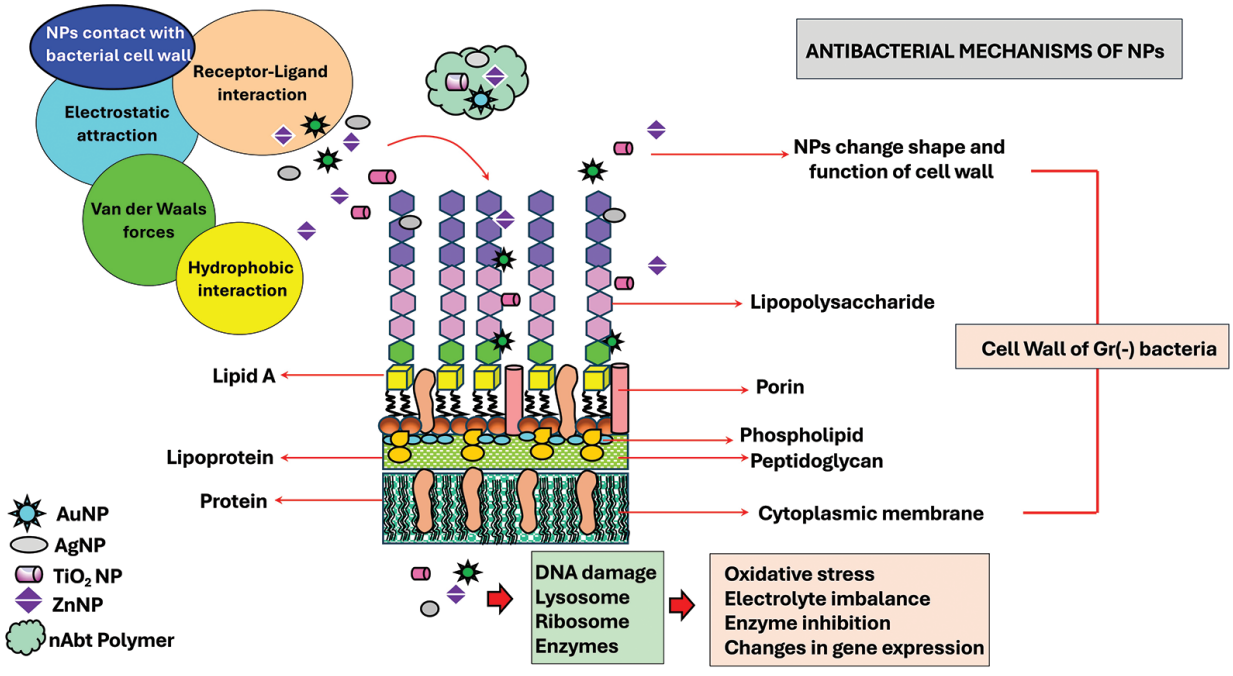

Figure 4. Nanoparticles (NPs) are required to be in contact with bacterial cells to achieve their antibacterial activity. Recognized modes of interaction between bacteria and nanoantibiotics (nAbts) include electrostatic attraction, van der Waals forces, receptor-ligand interactions, and hydrophobic interactions. Afterwards, NPs cross the bacterial membrane, assemble along metabolic pathways, and affect the structure and function of the cell membrane. Once inside the bacterial cell, NPs interact with essential components like DNA, lysosomes, ribosomes, and enzymes, causing oxidative stress, heterogeneous alterations, modified cell membrane permeability, electrolyte imbalances, enzyme inhibition, protein deactivation, as well as alteration in gene expression.

Ag: Silver, Au: Gold, Gr: Gram, TiO_2: Titanium dioxide, Zn: Zinc.

agents. Thereby they improve the killing potential of antibiotics or antimicrobial peptides, if they are administered in combination against pre-formed biofilms. Ciprofloxacin-loaded poly (lactic-co-glycolic acid) NPs, functionalized with DNase, exhibit antibiofilm activity *against P. aeruginosa* biofilms. Apart from releasing ciprofloxacin in a controlled manner, those NPs are capable of disassembling the biofilm by degrading the eDNA that stabilize the biofilm matrix (104). Drug delivery by lipid or polymer NPs is used to overcome biofilm resistance. The fusogenic liposomes maintain the release of antimicrobial agents in close proximity to the bacterial cells. These nanocarriers protect the antimicrobial compound from deleterious interactions with the matrix, enzymatic inactivation, and degradation by various bacterial and host components at the targeted infection site. Thereby, these nAbts are readily penetrate the biofilm (105). Anti-biofilm peptides have broad-spectrum activity against both Gram-positive and Gram-negative bacteria within biofilms. Furthermore, they present vigorous synergy with conventional antibiotics (106). The polyethylene glycol shell, combined with its surface-adaptive charge, allows the mixed-shell polymeric micelles (MSPMs) to penetrate and accumulate extensively in *Staphylococcal* biofilms. After the binding of micellar nanocarriers to the *Staphylococcal* cell surface, bacterial enzymes degrade the MSPM core, and thus the nAbt releases its antimicrobial content resulting in the killing of bacteria in the deep layers of a biofilm. This mechanism creates an extremely effective pathway to control blood-accessible *Staphylococcal* biofilms with antimicrobials, through bypassing biofilm recalcitrance to antimicrobial penetration (107) (Figure 5). Through this novel technology, NP-based carriers can deliver antibiotics with improved biofilm penetration, thereby combating bacterial resistance.

CONCLUSION

Since current European legislation does not include limits on antimicrobial concentrations in treated wastewater from WTPs, various antimicrobials have been detected at concentrations above acceptable levels in numerous WTPs across many European countries. Therefore, agricultural workers and their families, product traders, agricultural product consumers, residents of areas and eco-system irrigated with wastewater are exposed to wastewater containing high levels of ARGs (108). Hence, antimicrobial resistance is a significant public health problem worldwide. Given the important and interdependent aspects of this resistance affecting humans, animals, and the environment, a "One Health" approach is necessary to combat it. The One Health concept encompasses environmental transmission pathways and ecological selection pressures that sustain resistance beyond host organisms. Whole genome sequencing performed to determine the ARG profile and phylogenetic relationship in strains isolated from humans and eco-system has revealed a significant overlap between major antibiotics and ARGs (109). This finding indicates that the pharmaceutical industry and healthcare systems will continue to struggle against ARB strains. Reducing the spread of antibiotic resistance and alleviating its burden in humans are critical. Combating antibiotic resistance and preventing the transmission of ARB and ARG to humans require implementing control strategies in four areas: (1) Diminishing the colonization capacity of exogenous ARB; (2) improving human colonization resistance and reducing HGT of ARG; (3) reversing ARB antibiotic resistance (110), and (4) determining the permissible antibiotic concentrations in wastewater discharged from wastewater plants and legally regulating wastewater treatment systems accordingly. Although new antibiotics have significantly reduced the

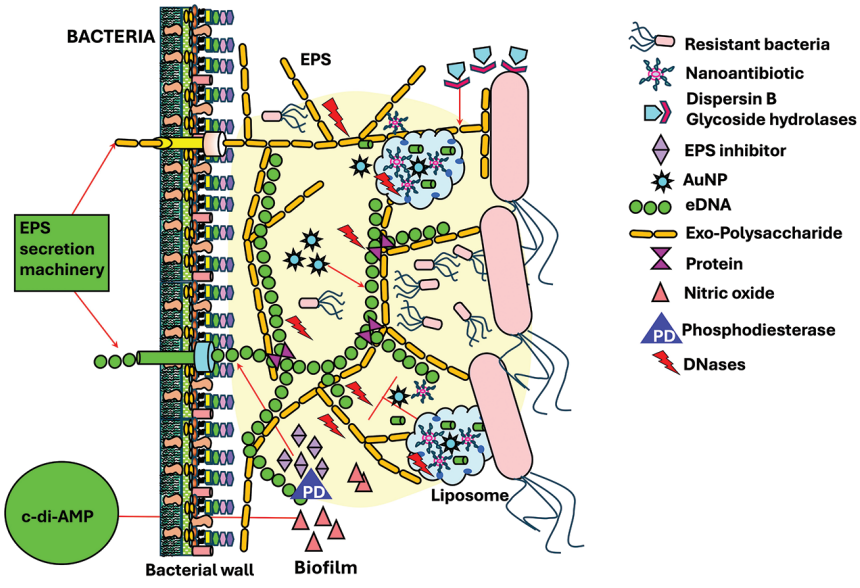

Figure 5. Bacterial biofilms, which have a structure rich in adhesine and extracellular matrix, protect microorganisms against conventional antibiotics. Nanoantibiotics (nAbts) target the extracellular polymeric substance (EPS) and decrease or completely remove the bacterial content of biofilm. Subsequent to the binding of micellar nanocarriers to bacterial cell surface, bacterial enzymes degrade mixed-shell-polymeric-micelles (MSPMs). Thus, nAbts penetrate and accumulate in bacterial biofilms. And then, antimicrobial compounds kill bacteria in the depth of a biofilm.

Au: Gold, EPS: Extracellular polymeric substance, c-di-AMP: Cyclic di-adenosine monophosphate, eDNA: Extracellular DNA, NP: Nanoparticle, PD: Phosphodiesterase.

persistence of bacterial infections, the development of resistance and biofilm formation generally reduce their effectiveness. Thus, newly synthesized antimicrobials lack the potential to provide a long-term solution. Combining antibiotics with NP-based carriers offers substantial advantages in overcoming antibiotic resistance, which is a global problem. Currently, only a small number of NP-based and hybrid nAbts have entered clinical practice. However, to address the problem of antibiotic resistance—which arises from the overuse of prophylactic or therapeutic antibiotics without appropriate medical indications, from the inappropriate selection of alternative antimicrobials, and from frequent switching between antimicrobial treatments—more clinical trials are needed to incorporate NP-based and hybrid nAbts into clinical practice.

Footnotes

Authorship Contributions

Concept: E.D.E., A.E., Design: E.D.E., A.E., Analysis or Interpretation: E.D.E., Literature Search: E.D.E., A.B.E., A.E., Writing: E.D.E., A.B.E., A.E.

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

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

REFERENCES

1. World Bank Group. Final report: drug-resistant infections a threat to our economic future [Internet]. Washington D.C.; 2017. Available from: https://documents1.worldbank.org/curated/en/323311493396993758/pdf/final-report.pdf
2. Iskandar K, Molinier L, Hallit S, Sartelli M, Catena F, Coccolini F, et al. Drivers of antibiotic resistance transmissionin low- and middle-income countriesfrom a “one health” perspective-a review. Antibiotics (Basel). 2020; 9: 372.
3. World Bank Group. Operational framework for strengthening human, animal, and environmental public health systems at their interface [Internet]. Washington D.C.; 2018. Available from: https://documents1.worldbank.org/curated/en/703711517234402168/pdf/123023-REVISED-PUBLIC-World-Bank-One-Health-Framework-2018.pdf
4. Laborda P, Sanz-García F, Ochoa-Sánchez LE, Gil-Gil T, Hernando-Amado S, Martínez JL. Wildlife and antibiotic resistance. Front Cell Infect Microbiol. 2022; 12: 873989.
5. Lester Y, Avisar D, Gozlan I, Mamane H. Removal of pharmaceuticals using combination of UV/H(2)O(2)/O(3)advanced oxidation process. Water Sci Technol. 2011; 64: 2230-8.
6. Jureczko M, Przystaś W. Ecotoxicity risk of presence of two cytostatic drugs: Bleomycin and vincristine and their binary mixture in aquatic environment. Ecotoxicol Environ Saf. 2019; 172: 210-5.
7. Białk-Bielińska A, Mulkiewicz E, Stokowski M, Stolte S, Stepnowski P. Acute aquatic toxicity assessment of six anti-cancer drugs and one metabolite using biotest battery - biological effects and stability under test conditions. Chemosphere. 2017; 189: 689-98.
8. Lekskulchai V. Quantitation of anticancer drugs - cyclophosphamide and ifosfamide in urine and water sewage samples by gas chromatography-mass spectrometry. Int J Occup Med Environ Health. 2016; 29: 815-22.
9. Grzesiuk M, Mielecki D, Pilżys T, Garbicz D, Marcinkowski M, Grzesiuk E. How cyclophosphamide at environmentally relevant concentration influences Daphnia magna life history and its proteome. PLoS One. 2018; 13: e0195366.
10. Lutterbeck CA, Kern DI, Machado ÊL, Kümmerer K. Evaluation of the toxic effects of four anti-cancer drugs in plant bioassays and its potency for screening in the context of waste water reuse for irrigation. Chemosphere. 2015; 135: 403-10.
11. WHO. Global research agenda for antimicrobial resistance in human health. Switzerland: antimicrobial resistance division World Health Organization; 2004.
12. Straub JO, Oldenkamp R, Pfister T, Häner A. Environmental risk assessment for the active pharmaceutical ingredient mycophenolic acid in european surface waters. Environ Toxicol Chem. 2019; 38: 2259-78.
13. Ashfaq M, Nawaz Khan K, Saif Ur Rehman M, Mustafa G, Faizan Nazar M, Sun Q, et al. Ecological risk assessment of pharmaceuticals in the receiving environment of pharmaceutical wastewater in Pakistan. Ecotoxicol Environ Saf. 2017; 136: 31-9.
14. Langford K, Thomas KV. Input of selected human pharmaceutical metabolites into the Norwegian aquatic environment. J Environ Monit. 2011; 13: 416-21.
15. Verlicchi P, Zambello E. Pharmaceuticals and personal care products in untreated and treated sewage sludge: occurrence and environmental risk in the case of application on soil - a critical review. Sci Total Environ. 2015; 538: 750-67.
16. Martínez Bueno MJ, Boillot C, Munaron D, Fenet H, Casellas C, Gómez E. Occurrence of venlafaxine residues and its metabolites in marine mussels at trace levels: development of analytical method and a monitoring program. Anal Bioanal Chem. 2014; 406: 601-10.
17. Klosterhaus SL, Grace R, Hamilton MC, Yee D. Method validation and reconnaissance of pharmaceuticals, personal care products, and alkylphenols in surface waters, sediments, and mussels in an urban estuary. Environ Int. 2013; 54: 92-9.
18. Gonzalez-Rey M, Bebianno MJ. Effects of non-steroidal anti-inflammatory drug (NSAID) diclofenac exposure in mussel Mytilus galloprovincialis. Aquat Toxicol. 2014; 148: 221-30.
19. Gonzalez-Rey M, Mattos JJ, Piazza CE, Bainy ACD, Bebianno MJ. Effects of active pharmaceutical ingredients mixtures in mussel Mytilus galloprovincialis. Aquat Toxicol. 2014; 153: 12-26.
20. Solé M, Shaw JP, Frickers PE, Readman JW, Hutchinson TH. Effects on feeding rate and biomarker responses of marine mussels experimentally exposed to propranolol and acetaminophen. Anal Bioanal Chem. 2010; 396: 649-56.
21. Fahmi A, Palin V, Zhong X, Yang YT, Watts S, Ashcroft DM, et al. Evaluation of the impact of COVID-19 pandemic on hospital admission related to common infections: risk prediction models to tackle antimicrobial resistance in primary care. PLoS One. 2024; 19: e0311515.
22. WHO. 10 global health issues to track in 2021 [Internet]. 2020 [cited 2025 Aug 20]. Available from: https://www.who.int/news-room/spotlight/10-global-health-issues-to-track-in-2021
23. Subramanya SH, Czyż DM, Acharya KP, Humphreys H. The potential impact of the COVID-19 pandemic on antimicrobial resistance and antibiotic stewardship. Virusdisease. 2021; 32: 330-7.
24. Schlüter A, Szczepanowski R, Pühler A, Top EM. Genomics of IncP-1 antibiotic resistance plasmids isolated from wastewater treatment plants provides evidence for a widely accessible drug resistance gene pool. FEMS Microbiol Rev. 2007; 31: 449-77.
25. Liang X, Chen B, Nie X, Shi Z, Huang X, Li X. The distribution and partitioning of common antibiotics in water and sediment of the Pearl River Estuary, South China. Chemosphere. 2013; 92: 1410-6.
26. Li S, Shi W, Li H, Xu N, Zhang R, Chen X, et al. Antibiotics in water and sediments of rivers and coastal area of Zhuhai City, Pearl River estuary, south China. Sci Total Environ. 2018; 636: 1009-19.

27. Bai Y, Meng W, Xu J, Zhang Y, Guo C. Occurrence, distribution and bioaccumulation of antibiotics in the Liao River Basin in China. Environ Sci Process Impacts. 2014; 16: 586-93.
28. Butnariu M. Chapter 2 - Heavy metals as pollutants in the aquatic Black Sea ecosystem. In: Dar GH, Bhat RA, Qadri H, Al-Ghamdy KM, Hakeem KR, editors. Bacterial fish diseases [Internet]. Academic Press; 2022 [cited 2026 Jun 28]. p. 31-57. Available from: https://www.sciencedirect.com/science/article/pii/B9780323856249000038 doi:10.1016/B978-0-323-85624-9.00003-8
29. Xi F. The enrofloxacin pollution control from fish to environment. Mar Pollut Bull. 2024; 199: 115923.
30. Kelly KR, Brooks BW. Global aquatic hazard assessment of ciprofloxacin: exceedances of antibiotic resistance development and ecotoxicological thresholds. Prog Mol Biol Transl Sci. 2018; 159: 59-77.
31. Wang K, Gao D, Xu J, Cai L, Cheng J, Yu Z, et al. Interaction of ciprofloxacin with the activated sludge of the sewage treatment plant. Environ Sci Pollut Res Int. 2018; 25: 35064-73.
32. Schafhauser BH, Kristofco LA, de Oliveira CMR, Brooks BW. Global review and analysis of erythromycin in the environment: occurrence, bioaccumulation and antibiotic resistance hazards. Environ Pollut. 2018; 238: 440-51.
33. Le Bris H, Pouliquen H. Experimental study on the bioaccumulation of oxytetracycline and oxolinic acid by the blue mussel (Mytilus edulis). An evaluation of its ability to bio-monitor antibiotics in the marine environment. Mar Pollut Bull. 2004; 48: 434-40.
34. Li W, Shi Y, Gao L, Liu J, Cai Y. Investigation of antibiotics in mollusks from coastal waters in the Bohai Sea of China. Environ Pollut. 2012; 162: 56-62.
35. He X, Wang Z, Nie X, Yang Y, Pan D, Leung AOW, et al. Residues of fluoroquinolones in marine aquaculture environment of the Pearl River Delta, South China. Environ Geochem Health. 2012; 34: 323-35.
36. Epelde L, Lanzén A, Blanco F, Urich T, Garbisu C. Adaptation of soil microbial community structure and function to chronic metal contamination at an abandoned Pb-Zn mine. FEMS Microbiol Ecol. 2015; 91: 1-11.
37. Engin AB, Engin ED, Engin A. Effects of co-selection of antibiotic-resistance and metal-resistance genes on antibiotic-resistance potency of environmental bacteria and related ecological risk factors. Environ Toxicol Pharmacol. 2023; 98: 104081.
38. Zhao Y, Cocerva T, Cox S, Tardif S, Su JQ, Zhu YG, et al. Evidence for co-selection of antibiotic resistance genes and mobile genetic elements in metal polluted urban soils. Sci Total Environ. 2019; 656: 512-20
39. Ma L, Xia Y, Li B, Yang Y, Li LG, Tiedje JM, et al. Metagenomic assembly reveals hosts of antibiotic resistance genes and the shared resistome in pig, chicken, and human feces. Environ Sci Technol. 2016; 50: 420-7.
40. Taylor NGH, Verner-Jeffreys DW, Baker-Austin C. Aquatic systems: maintaining, mixing and mobilising antimicrobial resistance? Trends Ecol Evol. 2011; 26: 278-84.
41. O'Neill J. Tackling drug-resistant infections globally: final report and recommendations [Internet]. 2016 May. p. 1-81. Available from: https://amr-review.org/sites/default/files/160518_Final%20paper_with%20cover.pdf
42. Wan X, Li Q, Li Z, Shi L, Pan Y, Li M, et al. MGE-associated ARGs exhibit higher expression efficiency than chromosomal non-MGE loci and predominantly contribute to resistance expression in pig farm wastewater. Environ Int. 2025; 202: 109653.
43. Laws M, Shaaban A, Rahman KM. Antibiotic resistance breakers: current approaches and future directions. FEMS Microbiol Rev. 2019; 43: 490-516.
44. Cycoń M, Mrozik A, Piotrowska-Seget Z. Antibiotics in the soil environment-degradation and their impact on microbial activity and diversity. Front Microbiol. 2019; 10: 338.
45. Le Page G, Gunnarsson L, Snape J, Tyler CR. Integrating human and environmental health in antibiotic risk assessment: a critical analysis of protection goals, species sensitivity and antimicrobial resistance. Environ Int. 2017; 109: 155-69.
46. Bengtsson-Palme J, Larsson DGJ. Concentrations of antibiotics predicted to select for resistant bacteria: proposed limits for environmental regulation. Environ Int. 2016; 86: 140-9.
47. Bouissou-Schurtz C, Houeto P, Guerbet M, Bachelot M, Casellas C, Mauclaire AC, et al. Ecological risk assessment of the presence of pharmaceutical residues in a French national water survey. Regul Toxicol Pharmacol. 2014; 69: 296-303.
48. Christou A, Agüera A, Bayona JM, Cytryn E, Fotopoulos V, Lambropoulou D, et al. The potential implications of reclaimed wastewater reuse for irrigation on the agricultural environment: the knowns and unknowns of the fate of antibiotics and antibiotic resistant bacteria and resistance genes - a review. Water Res. 2017; 123: 448-67.
49. He LX, He LY, Gao FZ, Wu DL, Ye P, Cheng YX, et al. Antibiotics, antibiotic resistance genes and microbial community in grouper mariculture. Sci Total Environ. 2022; 808: 152042.
50. Cabello FC, Godfrey HP, Tomova A, Ivanova L, Dölz H, Millanao A, et al. Antimicrobial use in aquaculture re-examined: its relevance to antimicrobial resistance and to animal and human health. Environ Microbiol. 2013; 15: 1917-42.
51. Zheng S, Qiu X, Chen B, Yu X, Liu Z, Zhong G, et al. Antibiotics pollution in Jiulong River estuary: source, distribution and bacterial resistance. Chemosphere. 2011; 84: 1677-85.
52. Blackburn JK, Mitchell MA, Blackburn MCH, Curtis A, Thompson BA. Evidence of antibiotic resistance in free-swimming, top-level marine predatory fishes. J Zoo Wildl Med. 2010; 41: 7-16.
53. Bonnedahl J, Drobni M, Gauthier-Clerc M, Hernandez J, Granholm S, Kayser Y, et al. Dissemination of Escherichia coli with CTX-M type ESBL between humans and yellow-legged gulls in the south of France. PLoS One. 2009; 4: e5958.
54. Moore JE, Rao JR, Moore PJA, Millar BC, Goldsmith CE, Loughrey A, et al. Determination of total antibiotic resistance in waterborne bacteria in rivers and streams in Northern Ireland: can antibiotic-resistant bacteria be an indicator of ecological change? Aquat Ecol. 2010; 44: 349-58.
55. Poirel L, Madec JY, Lupo A, Schink AK, Kieffer N, Nordmann P, et al. Antimicrobial resistance in Escherichia coli. Microbiol Spectr. 2018; 6.
56. Kvesić M, Kalinić H, Dželalija M, Šamanić I, Andričević R, Maravić A. Microbiome and antibiotic resistance profiling in submarine effluent-receiving coastal waters in Croatia. Environ Pollut. 2022; 292: 118282.
57. Xu N, Qiu D, Zhang Z, Wang Y, Chen B, Zhang Q, et al. A global atlas of marine antibiotic resistance genes and their expression. Water Res. 2023; 244: 120488.
58. Abe K, Nomura N, Suzuki S. Biofilms: hot spots of horizontal gene transfer (HGT) in aquatic environments, with a focus on a new HGT mechanism. FEMS Microbiol Ecol. 2020; 96: fiaa031.
59. GBD 2021 Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance 1990-2021: a systematic analysis with forecasts to 2050. Lancet. 2024; 404: 1199-226.

60. Lerminiaux NA, Cameron ADS. Horizontal transfer of antibiotic resistance genes in clinical environments. Can J Microbiol. 2019; 65: 34-44.
61. Biller SJ, Schubotz F, Roggensack SE, Thompson AW, Summons RE, Chisholm SW. Bacterial vesicles in marine ecosystems. Science. 2014; 343: 183-6.
62. von Wintersdorff CJH, Penders J, van Niekerk JM, Mills ND, Majumder S, van Alphen LB, et al. Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. Front Microbiol. 2016; 7: 173.
63. Li B, Qiu Y, Song Y, Lin H, Yin H. Dissecting horizontal and vertical gene transfer of antibiotic resistance plasmid in bacterial community using microfluidics. Environ Int. 2019; 131: 105007.
64. Wolska KI. Horizontal DNA transfer between bacteria in the environment. Acta Microbiol Pol. 2003; 52: 233-43.
65. Li W, Zhang G. Detection and various environmental factors of antibiotic resistance gene horizontal transfer. Environ Res. 2022; 212: 113267.
66. McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB, Paul JH. High frequency of horizontal gene transfer in the oceans. Science. 2010; 330: 50.
67. Rodríguez-Beltrán J, DelaFuente J, León-Sampedro R, MacLean RC, San Millán Á. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. Nat Rev Microbiol. 2021; 19: 347-59.
68. Shintani M, Vestergaard G, Milaković M, Kublik S, Smalla K, Schloter M, et al. Integrons, transposons and IS elements promote diversification of multidrug resistance plasmids and adaptation of their hosts to antibiotic pollutants from pharmaceutical companies. Environ Microbiol. 2023; 25: 3035-51.
69. Li J, Zuo J, Xu H, Yang J, Hu Y, Han Y, et al. Sub-inhibitory gentamicin promotes extracellular vesicles biogenesis and blaNDM dissemination in carbapenem-resistant Escherichia coli via mrdA/mrdB pathway. Vet Microbiol. 2025; 310: 110704.
70. Martínez MMB, Bonomo RA, Vila AJ, Maffía PC, González LJ. On the offensive: the role of outer membrane vesicles in the successful dissemination of New Delhi metallo-β-lactamase (NDM-1). mBio. 2021; 12: e0183621.
71. González LJ, Bahr G, Nakashige TG, Nolan EM, Bonomo RA, Vila AJ. Membrane anchoring stabilizes and favors secretion of New Delhi metallo-β-lactamase. Nat Chem Biol. 2016; 12: 516-22.
72. Aljohni MS, Harun-Ur-Rashid M, Selim S. Emerging threats: antimicrobial resistance in extended-spectrum beta-lactamase and carbapenem-resistant Escherichia coli. Microb Pathog. 2025; 200: 107275.
73. Wang X, Zhang H, Yu S, Li D, Gillings MR, Ren H, et al. Inter-plasmid transfer of antibiotic resistance genes accelerates antibiotic resistance in bacterial pathogens. ISME J. 2024; 18: wrad032.
74. Slizovskiy IB, Mukherjee K, Dean CJ, Boucher C, Noyes NR. Mobilization of antibiotic resistance: are current approaches for colocalizing resistomes and mobilomes useful? Front Microbiol. 2020; 11: 1376.
75. Cheng J, Wang Q, Chen Y, Ye Y, Li H, Li X, et al. Phenotypic and molecular characterization of a novel beta-lactamase carried by Klebsiella pneumoniae, CTX-M-72, derived from CTX-M-3. J Gen Appl Microbiol. 2009; 55: 207-16.
76. Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MOA, Dantas G. The shared antibiotic resistome of soil bacteria and human pathogens. Science. 2012; 337: 1107-11.
77. Fernandez-Moure JS, Evangelopoulos M, Colvill K, Van Eps JL, Tasciotti E. Nanoantibiotics: a new paradigm for the treatment of surgical infection. Nanomedicine (Lond). 2017; 12: 1319-34.
78. Edson JA, Kwon YJ. Design, challenge, and promise of stimuli-responsive nanoantibiotics. Nano Converg. 2016; 3: 26.
79. Engin AB, Engin A. Nanoantibiotics: a novel rational approach to antibiotic resistant infections. Curr Drug Metab. 2019; 20: 720-41.
80. Ndugire W, Raviranga NGH, Lao J, Ramström O, Yan M. Gold nanoclusters as nanoantibiotic auranofin analogues. Adv Healthc Mater. 2022; 11: e2101032.
81. Zheng Y, Liu W, Chen Y, Li C, Jiang H, Wang X. Conjugating gold nanoclusters and antimicrobial peptides: from aggregation-induced emission to antibacterial synergy. J Colloid Interface Sci. 2019; 546: 1-10.
82. Duan F, Feng X, Jin Y, Liu D, Yang X, Zhou G, et al. Metal-carbenicillin framework-based nanoantibiotics with enhanced penetration and highly efficient inhibition of MRSA. Biomaterials. 2017; 144: 155-65.
83. Chakraborty SP, Sahu SK, Mahapatra SK, Santra S, Bal M, Roy S, et al. Nanoconjugated vancomycin: new opportunities for the development of anti-VRSA agents. Nanotechnology. 2010; 21: 105103.
84. Chamundeeswari M, Sobhana SSL, Jacob JP, Kumar MG, Devi MP, Sastry TP, et al. Preparation, characterization and evaluation of a biopolymeric gold nanocomposite with antimicrobial activity. Biotechnol Appl Biochem. 2010; 55: 29-35.
85. Brown AN, Smith K, Samuels TA, Lu J, Obare SO, Scott ME. Nanoparticles functionalized with ampicillin destroy multiple-antibiotic-resistant isolates of Pseudomonas aeruginosa and Enterobacter aerogenes and methicillin-resistant Staphylococcus aureus. Appl Environ Microbiol. 2012; 78: 2768-74.
86. Lai HZ, Chen WY, Wu CY, Chen YC. Potent antibacterial nanoparticles for pathogenic bacteria. ACS Appl Mater Interfaces. 2015; 7: 2046-54.
87. Fakhri A, Tahami S, Naji M. Synthesis and characterization of core-shell bimetallic nanoparticles for synergistic antimicrobial effect studies in combination with doxycycline on burn specific pathogens. J Photochem Photobiol B. 2017; 169: 21-6.
88. Assali M, Zaid AN, Abdallah F, Almasri M, Khayyat R. Single-walled carbon nanotubes-ciprofloxacin nanoantibiotic: strategy to improve ciprofloxacin antibacterial activity. Int J Nanomedicine. 2017; 12: 6647-59.
89. Canaparo R, Foglietta F, Giuntini F, Pepa CD, Dosio F, Serpe L. Recent developments in antibacterial therapy: focus on stimuli-responsive drug-delivery systems and therapeutic nanoparticles. Molecules. 2019; 24: 1991.
90. de la Rica R, Aili D, Stevens MM. Enzyme-responsive nanoparticles for drug release and diagnostics. Adv Drug Deliv Rev. 2012; 64: 967-78.
91. Wang YL, He M, Miron RJ, Chen AY, Zhao YB, Zhang YF. Temperature/pH-sensitive nanoantibiotics and their sequential assembly for optimal collaborations between antibacterial and immunoregulation. ACS Appl Mater Interfaces. 2017; 9: 31589-99.
92. Wolfram J, Zhu M, Yang Y, Shen J, Gentile E, Paolino D, et al. Safety of nanoparticles in medicine. Curr Drug Targets. 2015; 16: 1671-81.
93. Neagu M, Piperigkou Z, Karamanou K, Engin AB, Docea AO, Constantin C, et al. Protein bio-corona: critical issue in immune nanotoxicology. Arch Toxicol. 2017; 91: 1031-48.
94. Cedervall T, Lynch I, Lindman S, Berggård T, Thulin E, Nilsson H, et al. Understanding the nanoparticle-protein corona using methods to quantify exchange rates and affinities of proteins for nanoparticles. Proc Natl Acad Sci USA. 2007; 104: 2050-5.
95. Ge C, Du J, Zhao L, Wang L, Liu Y, Li D, et al. Binding of blood proteins to carbon nanotubes reduces cytotoxicity. Proc Natl Acad Sci USA. 2011; 108: 16968-73.

96. Dobrovolskaia MA, McNeil SE. Immunological properties of engineered nanomaterials. Nat Nanotechnol. 2007; 2: 469-78.
97. Gref R, Minamitake Y, Peracchia MT, Trubetskoy V, Torchilin V, Langer R. Biodegradable long-circulating polymeric nanospheres. Science. 1994; 263: 1600-3.
98. Mouton JW, Ambrose PG, Canton R, Drusano GL, Harbarth S, MacGowan A, et al. Conserving antibiotics for the future: new ways to use old and new drugs from a pharmacokinetic and pharmacodynamic perspective. Drug Resist Updat. 2011; 14: 107-17.
99. Muzammil S, Hayat S, Fakhar-E-Alam M, Aslam B, Siddique MH, Nisar MA, et al. Nanoantibiotics: Future nanotechnologies to combat antibiotic resistance. Front Biosci (Elite Ed). 2018; 10: 352-74.
100. Kalhapure RS, Suleman N, Mocktar C, Seedat N, Govender T. Nanoengineered drug delivery systems for enhancing antibiotic therapy. J Pharm Sci. 2015; 104: 872-905.
101. Huh AJ, Kwon YJ. "Nanoantibiotics": a new paradigm for treating infectious diseases using nanomaterials in the antibiotics resistant era. J Control Release. 2011; 156: 128-45.
102. Liu JL, Zhang WJ, Li XD, Yang N, Pan WS, Kong J, et al. Sustained-release genistein from nanostructured lipid carrier suppresses human lens epithelial cell growth. Int J Ophthalmol. 2016; 9: 643-9.
103. Okshevsky M, Regina VR, Meyer RL. Extracellular DNA as a target for biofilm control. Curr Opin Biotechnol. 2015; 33: 73-80.
104. Baelo A, Levato R, Julián E, Crespo A, Astola J, Gavaldà J, et al. Disassembling bacterial extracellular matrix with DNase-coated nanoparticles to enhance antibiotic delivery in biofilm infections. J Control Release. 2015; 209: 150-8.
105. Forier K, Raemdonck K, De Smedt SC, Demeester J, Coenye T, Braeckmans K. Lipid and polymer nanoparticles for drug delivery to bacterial biofilms. J Control Release. 2014; 190: 607-23.
106. Pletzer D, Coleman SR, Hancock RE. Anti-biofilm peptides as a new weapon in antimicrobial warfare. Curr Opin Microbiol. 2016; 33: 35-40.
107. Liu Y, Busscher HJ, Zhao B, Li Y, Zhang Z, van der Mei HC, et al. Surface-adaptive, antimicrobially loaded, micellar nanocarriers with enhanced penetration and killing efficiency in staphylococcal biofilms. ACS Nano. 2016; 10: 4779-89.
108. Adegoke AA, Amoah ID, Stenström TA, Verbyla ME, Mihelcic JR. Epidemiological evidence and health risks associated with agricultural reuse of partially treated and untreated wastewater: a review. Front Public Health. 2018; 6: 337.
109. Li C, Lee LHA, Nanayakkara Sapugahawatte D, Lee WS, Cheung IYY, Chow VCY, et al. One health approach to characterize antimicrobial resistance in *Morganella morganii* in humans, freshwater fish and pigs. JAC Antimicrob Resist. 2026; 8: dlag064.
110. Ding D, Wang B, Zhang X, Zhang J, Zhang H, Liu X, et al. The spread of antibiotic resistance to humans and potential protection strategies. Ecotoxicol Environ Saf. 2023; 254: 114734.

Gazi Med J 2026;37(4):596-604

DOI: http://dx.doi.org/10.12996/gmj.2026.4723

Artificial Intelligence in Peroral Endoscopic Myotomy: Current Evidence, Limitations, and Future Directions

Peroral Endoskopik Miyotomide Yapay Zeka: Güncel Kanıtlar, Kısıtlılıklar ve Geleceğe Yönelik Yaklaşımlar

Ahmed Abdallah Salman[1], Ahmed Abdallah[2]

[1]Department of Internal Medicine, Cairo University, Faculty of Medicine, Cairo, Egypt
[2]Department of General Surgery, Cairo University, Faculty of Medicine, Cairo, Egypt

ABSTRACT

Peroral endoscopic myotomy (POEM) is an established treatment for achalasia and selected spastic esophageal motility disorders, but several aspects of care remain highly operator-dependent. These include early diagnosis, patient selection, procedural planning, myotomy length, intra-procedural safety, training assessment, and prediction of post-procedure reflux. Artificial intelligence (AI) may help address these limitations by integrating endoscopic video, high-resolution manometry, timed barium esophagram, EndoFLIP, clinical variables, procedural data, and follow-up outcomes into more personalized decision-support tools. This narrative review critically examines the evidence for AI across the POEM pathway and distinguishes direct achalasia or POEM evidence from adjacent and conceptual applications. Early AI applications are most likely to emerge in achalasia detection during diagnostic endoscopy, automated motility interpretation, video-based procedural assessment, and prediction of clinical response or post-POEM reflux. More ambitious applications, such as real-time tunnel orientation, vessel recognition, mucosal injury alerts, and personalized myotomy planning, remain largely investigational. Most AI-specific POEM applications remain at an early or conceptual stage and require multicenter validation, standardized outcomes, and careful clinical governance. AI should support, rather than replace, expert judgment, and an AI-derived reflux estimate cannot replace standard follow-up or objective reflux assessment.

ÖZ

Peroral endoskopik miyotomi (POEM), akalazya ve seçilmiş spastik özofagus motilite bozukluklarının tedavisinde yerleşik bir yöntem olmakla birlikte, tedavi sürecinin çeşitli aşamaları halen büyük ölçüde uygulayıcı deneyimine bağlıdır. Bu aşamalar; erken tanı, hasta seçimi, işlem planlaması, miyotomi uzunluğunun belirlenmesi, işlem sırasında güvenliğin sağlanması, eğitim düzeyinin değerlendirilmesi ve işlem sonrası reflü gelişiminin öngörülmesini kapsamaktadır. Yapay zeka (YZ); endoskopik video görüntüleri, yüksek çözünürlüklü manometri, zamanlı baryumlu özofagus grafisi, EndoFLIP, klinik değişkenler, işleme ilişkin veriler ve takip sonuçlarını bütünleştirerek daha kişiselleştirilmiş karar destek araçlarının geliştirilmesine katkı sağlayabilir. Bu anlatısal derlemede, POEM sürecinin farklı aşamalarında YZ kullanımına ilişkin kanıtlar eleştirel olarak incelenmekte ve doğrudan akalazya veya POEM'e ilişkin kanıtlar, komşu alanlardaki ve kavramsal düzeydeki uygulamalardan ayrılmaktadır. YZ'nin erken dönemdeki uygulamalarının; tanısal endoskopi sırasında akalazyanın saptanması, motilite bulgularının otomatik olarak yorumlanması, video tabanlı işlem değerlendirmesi ve klinik yanıtın veya POEM sonrası reflünün öngörülmesi alanlarında ortaya çıkması beklenmektedir. Tünel yönünün gerçek zamanlı olarak belirlenmesi, damarların tanınması, mukozal yaralanma uyarılarının oluşturulması ve kişiselleştirilmiş miyotomi planlaması gibi daha ileri uygulamalar ise büyük ölçüde araştırma aşamasındadır. POEM'e özgü YZ uygulamalarının çoğu halen erken

Cite this article as: Salman AA, Abdallah A. Artificial intelligence in peroral endoscopic myotomy: current evidence, limitations, and future directions. Gazi Med J. 2026;37(4):596-604

Address for Correspondence/Yazışma Adresi: Ahmed Abdallah Salman, Department of Internal Medicine, Cairo University, Faculty of Medicine, Cairo, Egypt
E-mail / E-posta: Awea84@kasralainy.edu.eg, Awea844@gmail.com
ORCID ID: orcid.org/0009-0006-4244-9377

Received/Geliş Tarihi: 24.05.2026
Accepted/Kabul Tarihi: 16.08.2026
Publication Date/Yayınlanma Tarihi: 15.10.2026

©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

ÖZ

veya kavramsal aşamadadır ve çok merkezli doğrulama çalışmalarına, standardize edilmiş sonuç ölçütlerine ve dikkatli bir klinik yönetişim yaklaşımına ihtiyaç duymaktadır. YZ, uzman klinik değerlendirmesinin yerine geçmekten ziyade onu desteklemelidir. Ayrıca YZ aracılığıyla elde edilen reflü tahmini, standart takip sürecinin veya objektif reflü değerlendirmesinin yerini alamaz.

Keywords: Artificial intelligence, peroral endoscopic myotomy, achalasia, computer vision, high-resolution manometry, gastroesophageal reflux disease

Anahtar Sözcükler: Yapay zeka, peroral endoskopik miyotomi, akalazya, bilgisayarlı görü, yüksek çözünürlüklü manometri, gastroözofageal reflü hastalığı

INTRODUCTION

POEM as a Platform for Precision Endoscopy

Peroral endoscopic myotomy (POEM) has transformed the management of achalasia and selected spastic esophageal motility disorders. Since its introduction as a natural-orifice endoscopic alternative to laparoscopic Heller myotomy, POEM has moved from an experimental third-space endoscopy technique to an established therapeutic option for idiopathic achalasia, type III achalasia, distal esophageal spasm, jackhammer esophagus, and selected cases of failed prior therapy (1,2). Its main appeal lies in its ability to provide a flexible, endoluminal myotomy without external incisions, while allowing the endoscopist to tailor the length and orientation of the myotomy according to the underlying motility pattern.

Contemporary guidelines and reviews support POEM as a safe and effective treatment for achalasia, with outcomes comparable to established alternatives such as laparoscopic Heller myotomy and generally superior efficacy to pneumatic dilation in many clinical settings (2-4). However, POEM remains a highly operator-dependent procedure. Important decisions are still based on endoscopist experience, center practice, and interpretation of heterogeneous pre-procedural data. These include whether POEM is the best treatment for a given patient, how long the esophageal and gastric myotomy should be, whether an anterior or posterior approach is preferable, how to manage sigmoid or advanced achalasia, and how to balance symptom relief against the risk of post-POEM gastroesophageal reflux disease (GERD) (2,4).

This shift creates a strong rationale for artificial intelligence (AI). In achalasia care, clinicians already integrate multiple data streams, including symptoms, endoscopic findings, high-resolution manometry (HRM), timed barium esophagram (TBE), functional lumen imaging probe (FLIP; commonly performed with the EndoFLIP system) measurements, prior interventions, anatomical complexity, and patient-specific reflux risk. However, these variables are rarely combined in a systematic or quantitative way. AI may help convert this fragmented information into clinically useful predictions, supporting earlier diagnosis, improved patient selection, personalized procedural planning, objective intra-procedural guidance, and tailored follow-up (5). In this review, more precise terms such as machine learning, clinical prediction model, deep learning, computer vision, or transformer model are used when the specific method is known.

The earliest AI applications relevant to POEM are likely to emerge before and after the procedure rather than during the procedure itself. For example, computer vision models may assist in recognizing subtle endoscopic features of achalasia, while clinical prediction models may help estimate the probability of symptom response, recurrence, or post-POEM reflux. A recent proof-of-concept study showed that a model using upper endoscopy videos could detect achalasia with diagnostic accuracy comparable to expert endoscopists, supporting the concept that routine endoscopic video contains clinically useful information that may be under-recognized during standard visual assessment (6).

Nevertheless, AI in POEM remains at an early stage. Most available work relates to achalasia diagnosis, general AI in endoscopy, or outcome prediction rather than real-time AI-guided myotomy. Therefore, the purpose of this review is not to claim that AI is ready to perform or direct POEM, but to critically examine where it may add value across the POEM pathway. The central argument is that POEM is an ideal model for precision therapeutic endoscopy: it is technically complex, data-rich, outcome-driven, and vulnerable to variation in decision-making. AI may eventually help standardize this pathway while preserving the endoscopist's central role in judgment, technique, and patient-centered care.

Methods: Narrative Search Strategy and Evidence Classification

PubMed, MEDLINE, Embase, Scopus, Web of Science Core Collection, and Google Scholar were searched from database inception through April 2026. Search concepts combined "achalasia," "peroral endoscopic myotomy" or "POEM" with "artificial intelligence," "machine learning," "deep learning," "computer vision," "clinical prediction model," "risk score," "nomogram," "technical difficulty," "adverse event," "complication," "clinical failure," "poor responder," "high-resolution manometry," "functional lumen imaging probe," "FLIP," "EndoFLIP," "outcome prediction," "reflux," "training," or "video assessment." Relevant studies published through the April 2026 cutoff were considered, and reference lists of relevant guidelines, reviews, and eligible studies were also examined.

English-language, peer-reviewed human studies directly addressing achalasia or POEM were prioritized. Conventional regression-based POEM risk scores were included when they addressed treatment response, technical difficulty, adverse events, or reflux because they are methodological precursors to AI-enabled prediction; they were described separately and were not treated as deployable AI systems. Studies from endoscopic submucosal dissection (ESD), minimally invasive surgery, or general endoscopy were included only when they addressed an analogous video or workflow task and were labeled

as adjacent evidence. Reports were selected purposively for clinical relevance and methodological detail. This was a narrative review; duplicate independent screening, protocol registration, formal risk-of-bias assessment, and meta-analysis were not undertaken.

Evidence was classified as clinically evaluated when a complete system or conventional prediction model had been tested in a patient cohort, reader study, or clinical prediction cohort; proof of concept when evaluation was image-level, preclinical, a small feasibility study, or limited to a component task rather than a complete clinical model; adjacent when evidence came from a different procedure; and conceptual when no direct POEM evaluation was available. Retrospective design and internal validation were recorded as limitations rather than overriding the clinically evaluated label for completed prediction models. Conventional regression-based scores were identified explicitly so that clinical evaluation was not mistaken for AI-specific integration or readiness for routine care. These labels describe the maturity of evaluation and do not imply deployability.

Current Evidence Base and Maturity

The strongest direct evidence concerns recognition of achalasia during diagnostic endoscopy. Kaklamanos et al. (6) assembled 1,229 videos from 1,203 patients investigated for dysphagia at one center. Their transformer model used 1,036 videos from 1,010 patients for training, 98 patients and videos for validation, and 95 patients and videos for testing, with HRM as the reference standard. In the held-out test set, model accuracy was 0.926, F1 score 0.821, sensitivity 0.800, and negative predictive value 0.947; two expert endoscopists had accuracy 0.905, F1 score 0.710, sensitivity 0.550, and negative predictive value 0.893 (6). The single-center, enriched dysphagia cohort, exclusions, and small number of achalasia cases in the test set limit generalizability to routine unselected endoscopy.

Shiwaku et al. (7) conducted a multicenter reader study involving 83 endoscopists from 27 centers. Readers assessed 50 selected videos, including 25 with achalasia, before and after computer-aided detection. Sensitivity increased from 74.2% to 91.2% overall and from 66.9% to 91.9% among less-experienced readers (7). The artificial disease prevalence, repeated case set, and video-reading design do not establish performance in a live workflow. Both diagnostic studies are vulnerable to spectrum and selection bias when results from enriched cohorts and selected videos are extrapolated to routine endoscopy.

Evidence for automated physiological testing is less mature. HRM remains the diagnostic reference standard under Chicago Classification version 4.0 (8), and existing machine-learning work has mainly addressed component tasks rather than complete clinical interpretation (9). Czakó et al. (10) used single-center, markedly imbalanced image datasets and an internal image-level split without external patient-level validation of a complete motility diagnosis. Carlson et al. (11) studied FLIP panometry in 180 treatment-naive patients with achalasia, using 140 for training and 40 for testing. Test accuracy was 78% for spastic versus non-spastic achalasia and 55% for all three HRM subtypes, supporting a preliminary phenotype model rather than a replacement for HRM.

Conventional regression-based POEM prediction models provide direct clinical precursor evidence. Urakami et al. (12) developed a poor-response score in 244 patients and internally validated it by bootstrapping [area under the curve (AUC) 0.78]. Liu et al. (13) developed a major-adverse-event score in 3,135 patients with bootstrap internal validation (AUC 0.795), while a separate 3,385-patient study used a 1,693-patient training cohort and 1,692-patient internal validation cohort to develop a technical-difficulty nomogram (validation AUC 0.743) (14). These retrospective models are not deployable AI systems and lack independent prospective validation.

The principal direct machine-learning outcome-prediction study included 1,824 treatment-naive patients across multiple Japanese centers, 1,778 of whom underwent POEM. The area under the receiver operating characteristic curve was 0.70 for persistent symptoms and 0.61 for reflux esophagitis, indicating moderate and poor discrimination, respectively (15). No independent geographic validation or prospective decision-impact evaluation was reported. Figure 1 maps these direct findings alongside proof-of-concept, adjacent, and conceptual applications.

AI-assisted Diagnosis and Patient Selection Before POEM

Accurate diagnosis and careful patient selection are central to successful POEM. Although POEM is highly effective for achalasia, it is not appropriate for every patient with dysphagia or esophageal outflow obstruction. The decision to proceed with POEM requires integration of symptoms, endoscopic findings, HRM, TBE, FLIP data, previous interventions, anatomical complexity, patient preference, and reflux risk. Current diagnostic frameworks, including the Chicago Classification version 4.0, provide a structured approach to esophageal motility disorders, but interpretation can remain challenging in borderline cases, spastic disorders, opioid-associated dysmotility, and patients with prior therapy (8,16).

AI may improve the earliest stage of the POEM pathway by supporting recognition of achalasia during routine upper endoscopy. Achalasia can be missed because endoscopic findings are often subtle, including retained saliva or fluid, mild esophageal dilation, abnormal contractions, resistance at the esophagogastric junction (EGJ), or a tight-appearing lower esophageal sphincter. Recent computer-vision work suggests that AI can identify achalasia from upper endoscopy videos with performance comparable to expert endoscopists. This is clinically relevant because many patients experience delayed diagnosis and may receive repeated endoscopy or empirical reflux therapy before manometry confirms the disorder (6,7).

AI may also support interpretation of HRM, which remains the diagnostic cornerstone for achalasia and other esophageal motility disorders. Manometric interpretation requires specialist expertise and may be affected by artifacts, catheter position, patient tolerance, and interobserver variability. Machine-learning tools could help classify motility patterns, identify achalasia subtypes, distinguish true spastic disease from artifact, and flag complex studies for expert review. This is important for POEM because manometric phenotype influences procedural planning, particularly in type III achalasia and spastic disorders, where a longer proximal myotomy is often required (9,10).

Patient selection should extend beyond confirming the diagnosis. A 244-patient regression-based score used the pretreatment Eckardt score, previous treatment, sigmoid esophagus, and esophageal dilation grade to estimate poor response after POEM, but it was only

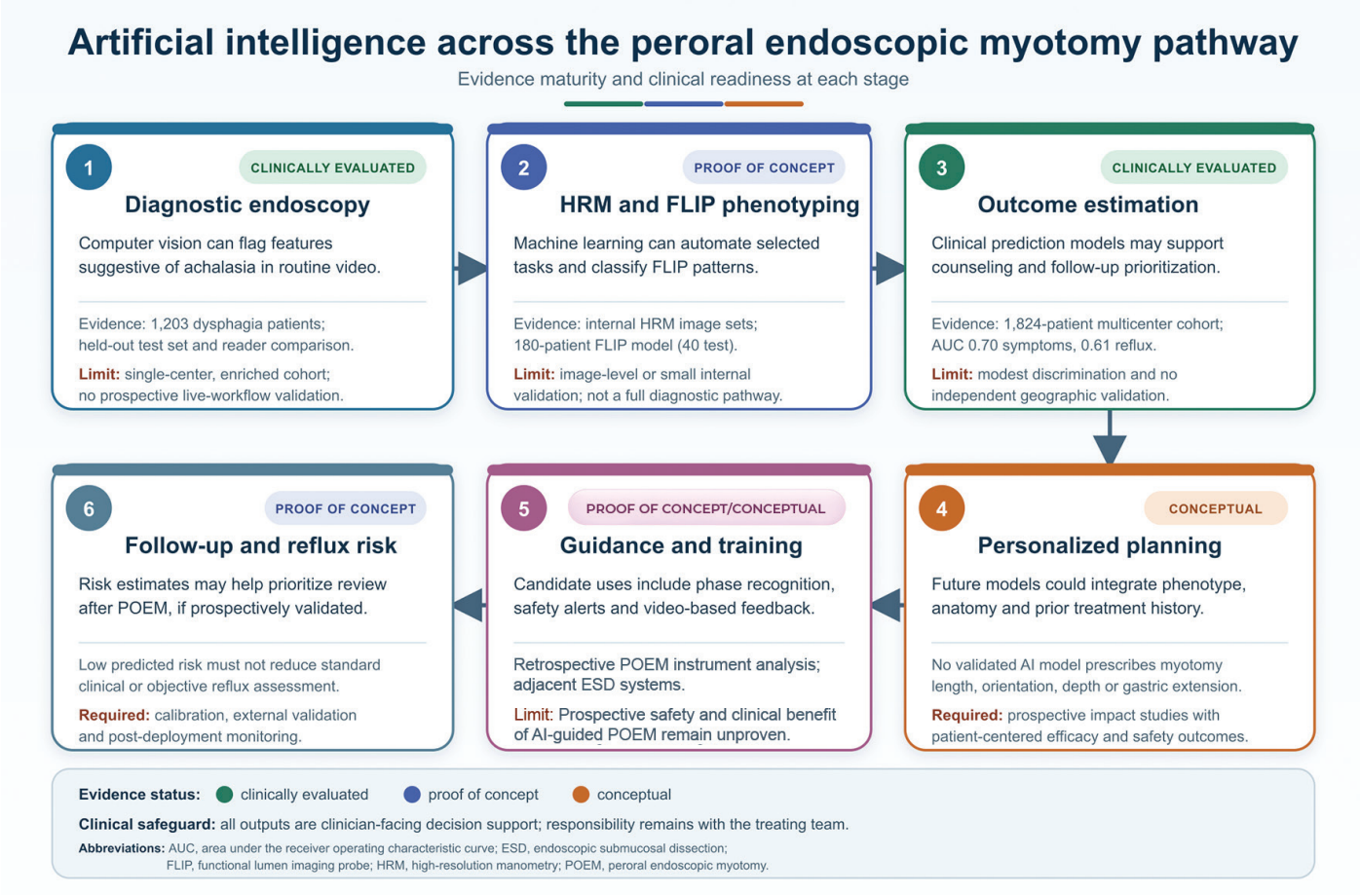

Figure 1. Evidence maturity across the proposed artificial intelligence-assisted peroral endoscopic myotomy pathway. "Clinically evaluated" denotes testing in a patient, reader, or prediction cohort and does not indicate readiness for routine care. Direct clinical evidence includes diagnostic endoscopy, retrospective machine-learning outcome estimation, and conventional regression-based POEM risk scores that are methodological precursors rather than deployable AI systems. Automated high-resolution manometry and functional lumen imaging probe interpretation remain proof of concept; personalized myotomy prescription, real-time intra-procedural guidance, and automated training assessment remain conceptual for peroral endoscopic myotomy. A low predicted reflux risk must not replace standard clinical follow-up or objective reflux assessment.

AI: Artificial intelligence, AUC: Area under the receiver operating characteristic curve, ESD: Endoscopic submucosal dissection, FLIP: Functional lumen imaging probe, HRM: High-resolution manometry, POEM: Peroral endoscopic myotomy. (This original figure was created by the authors for this review.)

internally validated and did not compare POEM with alternative treatments (12). A clinically useful model would ideally estimate the likely benefit of POEM compared with pneumatic dilation, laparoscopic Heller myotomy with fundoplication, botulinum toxin injection, redo myotomy, or conservative management. Contemporary guidelines support POEM as a safe and effective treatment for achalasia, with particular value in type III achalasia because the myotomy can be extended proximally according to the spastic segment. However, treatment selection remains complex in patients with advanced sigmoid achalasia, severe esophageal dilation, previous interventions, frailty, high reflux risk, or discordant symptoms (2,3,16).

Functional assessment may further strengthen AI-assisted selection. FLIP can assess EGJ distensibility and provide complementary information before, during, and after myotomy. Persistently low distensibility after treatment has been associated with poorer symptomatic outcomes, while excessive disruption of the antireflux barrier may increase reflux risk. Future clinical prediction models could combine distensibility metrics with HRM, TBE, endoscopic findings, and symptom scores to identify patients most likely to benefit from POEM and estimate the degree of myotomy needed to achieve adequate outflow relief without unnecessary disruption (11,17).

A practical AI-assisted pre-POEM workflow could include three levels. First, computer vision could flag suspected achalasia during diagnostic endoscopy. Second, automated analysis of manometry and functional testing could support disease phenotyping and identify complex or discordant cases. Third, clinical prediction models could estimate treatment response, technical difficulty, recurrence risk, and post-POEM reflux risk. Conventional regression scores already provide internally validated precursor evidence for several of these tasks, but they are not AI systems and have not shown prospective decision impact. An AI-assisted approach would not replace clinical judgment, but it could standardize decision-

making and reduce variation between centers, particularly where access to expert motility interpretation is limited (6,7,9-15).

At present, AI-specific pre-POEM assessment remains early and should be viewed as investigational rather than practice-changing. Most AI studies are proof of concept, retrospective, or focused on diagnosis rather than treatment selection; conventional POEM risk scores are clinically evaluated but remain internally validated precursors. The key question is not simply whether a model can classify achalasia, but whether it improves outcomes by helping clinicians select the right patient, intervention, and procedural strategy.

AI for Procedural Planning: Toward Personalized Myotomy

Procedural planning is one of the most attractive but least developed applications of AI in POEM. Although the procedure is often described in standard technical steps, the optimal strategy varies substantially between patients. The endoscopist must decide the orientation of the tunnel, the entry site, the length of esophageal and gastric myotomy, the depth of myotomy, and the degree of extension required across the EGJ. These decisions are usually based on endoscopic judgment, manometric subtype, anatomical features, and local expertise rather than objective prediction models (18).

The most important planning question is myotomy length. In type I and type II achalasia, shorter myotomy strategies may be adequate in selected patients, while type III achalasia and spastic esophageal disorders often require a longer myotomy that extends proximally to cover the spastic segment. This is a natural area for AI because a future model could integrate HRM, TBE, FLIP findings, endoscopic anatomy, symptom pattern, and prior treatment history to estimate the minimum effective myotomy length for each patient (19).

AI may also help reduce unnecessary procedural variation. A recent meta-analysis showed substantial heterogeneity in reported myotomy length across POEM studies, with longer pooled myotomy lengths in type III achalasia than in non-spastic achalasia. This variation reflects differences in disease phenotype, operator preference, and evolving procedural strategy. A validated prediction model could help distinguish patients who need extended myotomy from those in whom shorter myotomy may achieve adequate symptom relief while potentially reducing procedural time, tissue disruption, and reflux risk (20).

Functional assessment could make procedural planning more objective. FLIP provides real-time information on EGJ distensibility and may help assess whether the myotomy has achieved adequate physiological disruption. Studies have shown that distensibility changes during POEM can be measured incrementally and may correlate with clinical outcomes. In the future, a model could combine baseline and intra-procedural distensibility data with manometric phenotype to guide whether additional proximal or gastric extension is required (21,22).

Another planning issue is the risk of post-POEM reflux. Unlike laparoscopic Heller myotomy, POEM is not routinely combined with fundoplication, and reflux remains a major limitation. Proposed predictors remain plausible but heterogeneous candidates. In a 92-patient cohort, loss of inspiratory augmentation was associated with GERD in the evaluable HRM subgroup. Shorter intra-abdominal lower esophageal sphincter length and type III EGJ morphology showed nonsignificant trends; myotomy length and orientation were not significantly associated (23). These findings require external validation and cannot support routine model-based planning.

AI may also assist in predicting technical difficulty. Direct conventional evidence already exists: a regression nomogram developed in 3,385 patients and internally validated in 1,692 identified age, disease duration, sigmoid esophagus, mucosal edema, submucosal fibrosis, and tunnel length as predictors (14). This model is not an AI system and lacks independent prospective validation, but it establishes that technical-difficulty prediction is not entirely conceptual. Future models could integrate procedural videos, endoscopic images, radiology, and clinical variables to support operator, center, equipment, and backup-plan selection (18).

AI-guided prescription of myotomy length, orientation, depth, or gastric extension remains conceptual, whereas conventional models for technical difficulty and major adverse events have undergone internal clinical evaluation (13,14). No validated AI tool can prescribe a POEM strategy in routine practice. Future studies should develop clinically actionable models using multicenter datasets, standardized procedural variables, objective reflux testing, and long-term clinical outcomes, followed by external and prospective validation. The aim should not be to replace expert endoscopists, but to make planning more reproducible, transparent, and personalized (18-22).

AI for IIntra-procedural Guidance and Training

AI-assisted intra-procedural guidance represents the most ambitious future application in POEM. Unlike diagnostic endoscopy, POEM requires continuous recognition of dynamic procedural steps, including mucosal incision, submucosal tunnel creation, myotomy, hemostasis, extension across the EGJ, and mucosal closure. In theory, computer vision could provide real-time support by identifying the procedural phase, maintaining orientation within the tunnel, highlighting vessels, warning of mucosal injury, and prompting reassessment before closure. The main potential applications across the POEM pathway are summarized in Table 1.

The rationale for intra-procedural AI is strengthened by progress in adjacent therapeutic endoscopy. In ESD, computer-vision systems have been developed for workflow recognition, anatomical structure delineation, and real-time orientation support. Although these models cannot be directly transferred to POEM, they show that AI can analyze complex therapeutic endoscopy videos and provide cognitive assistance beyond simple lesion detection (24,25).

The supporting evidence remains preliminary and procedure-specific. Cao et al. (24) developed an ESD workflow-recognition system from 201,026 frames across 47 cases, performed a small external evaluation, and demonstrated real-time operation in an animal study. Scheppach et al. (25) trained anatomical segmentation on 1,011 still images from 30 ESDs and evaluated real-time use in 12 porcine procedures. These studies demonstrate technical feasibility in an adjacent procedure, not safety or benefit during human POEM.

A key intra-procedural target is recognition of the correct submucosal plane. Loss of orientation, difficult tunneling, fibrosis, bleeding, and inadvertent mucosal injury can increase technical difficulty and procedural risk. A model could be trained to distinguish the mucosal layer, submucosal tunnel, circular muscle fibers, longitudinal muscle, vessels, and areas of poor visibility. Zhu et al. (26) reported

Table 1. Evidence maturity of potential AI applications across the POEM pathway.

Stage of care	Potential AI application	Expected clinical value	Evidence status	Representative study or data type	Validation status	Current limitation
Diagnosis	Detection of achalasia from endoscopic images or videos	Earlier recognition and referral for manometry	Clinically evaluated	Kaklamanos et al. (6); Shiwaku et al. (7)	Single-center held-out test; multicenter reader study	Selected, enriched videos; no prospective live-workflow validation
Patient selection and risk stratification	Prediction of poor response, technical difficulty, major adverse events, and benefit from POEM	Better counseling, case selection, operator assignment, and procedural preparation	Clinically evaluated conventional prediction models; AI-specific integration and external validation remain unavailable	Urakami et al. (12); Liu et al. (13,14); Takahashi et al. (15)	Retrospective bootstrap, split-sample internal, or multicenter development evaluation	No independent prospective validation; conventional regression models are not deployable AI systems or comparative treatment selectors
Procedural planning	Prediction of optimal myotomy length, orientation, and gastric extension	More personalized and reproducible POEM strategy	Conceptual for AI-prescribed planning	Clinical POEM and FLIP cohorts; no direct AI planning model	No validated AI system	No evidence that an AI-prescribed plan improves efficacy or safety
Intra-procedural guidance	Instrument analysis, tunnel orientation, vessel detection and mucosal injury alerts	Improved safety and procedural consistency	Direct POEM instrument proof of concept; navigation/safety remain conceptual	Zhu et al. (26); adjacent ESD studies (24,25)	Retrospective POEM video analysis; no prospective safety/ benefit validation	Different anatomy and error costs; false-alert risk
Training	Automated video-based assessment of procedural steps and technical errors	Objective competency assessment and feedback	Automated competency scoring unvalidated	POEM assessment tool: 12 assessors and 10 videos (30)	Initial human-rating validation only	The assessment tool is not AI; automated scoring remains unvalidated
Follow-up	Prediction of clinical recurrence and post-POEM reflux	Tailored surveillance and acid-suppression strategy	Clinically evaluated prediction cohorts; no deployable AI system	Urakami et al. (12); Takahashi et al. (15); Kim et al. (23)	Retrospective internal or development-cohort evaluation; no decision-impact study	Heterogeneous predictors; low predicted risk cannot replace standard follow-up or objective assessment

AI: Artificial intelligence, ESD: Endoscopic submucosal dissection, FLIP: Functional lumen imaging probe, POEM: Peroral endoscopic myotomy.

retrospective instrument-based analysis of POEM videos; general surgical AI reviews provide adjacent evidence (27).

AI may also help standardize recognition and documentation of adverse events. Mucosal injury, bleeding, insufflation-related events, pain, and aspiration are recognized complications during or after POEM, although major adverse events are uncommon in experienced centers. A conventional regression score developed in 3,135 patients predicted major adverse events with bootstrap internal validation, but it was not an AI system and was not independently validated (13). Video-based models could automatically flag bleeding episodes, prolonged hemostasis, suspected mucosal breach, repeated loss of tunnel orientation, or difficult closure. Such systems must be evaluated against established prevention, recognition, and management standards before clinical use (28).

Training is likely to be a more realistic near-term application than real-time procedural control. POEM has a significant learning curve and requires both cognitive and technical skills, including patient selection, tunnel creation, controlled myotomy, hemostasis, complication recognition, and closure. Existing training guidance emphasizes structured curricula, expert supervision, simulation, animal models, and competency-based assessment rather than reliance on case numbers alone (29).

Recent work on video-based POEM assessment provides an important foundation for this direction. The POEM assessment tool was developed as a structured human rating instrument, not an AI system. In its initial validation, 12 assessors rated 10 procedural videos, with variable inter-rater reliability across domains (30). It may provide a competency framework for future automated assessment, but automated scoring has not yet been prospectively validated.

In the future, the most practical model may be a staged approach. Early tools could support retrospective video review and training assessment. More advanced systems could undergo silent-mode evaluation before carefully bounded phase-recognition or safety alerts are introduced. Only after robust prospective and human-factors validation should AI be considered for real-time decision support during active myotomy.

AI Prediction of Post-POEM Outcomes and Reflux

Prediction of post-POEM outcomes is clinically relevant because technical success does not always guarantee durable symptomatic

benefit. Most patients improve after POEM, but a minority develop persistent symptoms, recurrent dysphagia, need for retreatment, or clinically significant reflux. Conventional regression and machine-learning cohorts provide direct POEM-specific prediction evidence, although current models remain insufficient for routine decision-making (12,15). Outcome assessment relies on symptom scores, endoscopy, pH testing, TBE, HRM, and FLIP, but these data are often interpreted separately rather than integrated into a single personalized risk estimate.

AI could help predict clinical response by combining baseline and procedural variables. Relevant candidates include age, symptom duration, Eckardt score, achalasia subtype, esophageal diameter, sigmoid morphology, prior pneumatic dilation or Heller myotomy, lower esophageal sphincter pressure, integrated relaxation pressure, FLIP distensibility, myotomy length, gastric extension, and intra-procedural complications. Existing conventional and machine-learning models show that such estimation is feasible, but their internal or development-cohort performance is insufficient for individual treatment decisions without independent prospective validation (12,15).

Post-POEM reflux is particularly important because objective reflux can occur even when symptoms are mild or absent. Reported candidate predictors are heterogeneous rather than established. In a 92-patient retrospective cohort, loss of inspiratory augmentation was associated with GERD in the evaluable HRM subgroup; intra-abdominal lower esophageal sphincter length, EGJ morphology, myotomy length, and orientation were not significantly associated (23). Other cohorts have proposed different baseline and procedural variables, and differences in reflux definitions and testing further limit comparability. No reflux-prediction model is ready for routine use, and objective follow-up remains necessary (4,31).

The value of a reflux-prediction model would be greatest if it informed counseling and surveillance without weakening established care. A high predicted risk could support careful counseling, avoidance of excessive gastric extension, consideration of sling-fiber-preserving techniques or emerging antireflux POEM concepts, and structured post-procedure surveillance (32). Conversely, a low predicted risk must not be used to omit standard clinical follow-up, acid-suppression decisions, endoscopy, or objective reflux assessment when otherwise indicated.

Longitudinal prediction is another important opportunity. Some patients remain asymptomatic early after POEM but later develop recurrent symptoms, reflux esophagitis, abnormal acid exposure, or need for retreatment. A future model could update risk estimates using serial Eckardt scores, medication use, endoscopic findings, pH testing, TBE, and functional measurements. This could support risk-adapted follow-up, provided that standard surveillance is not withheld on the basis of an unvalidated estimate.

However, outcome prediction models must avoid overfitting and clinically misleading conclusions. Many POEM studies are retrospective, single-center, and heterogeneous in follow-up duration, reflux definitions, pH testing protocols, and outcome thresholds. Models trained on such data may perform well internally but fail when applied to different populations, operators, devices, and follow-up pathways. Future models should therefore use standardized definitions of clinical success, objective reflux testing, adequate follow-up, external validation, and transparent reporting and risk-of-bias standards (33,34).

The ideal AI-assisted POEM pathway would begin before the procedure and continue into follow-up. AI could flag suspected achalasia during endoscopy, support manometric phenotyping, predict benefit from POEM, assist with procedural planning, provide video-based quality assessment, and estimate the risk of recurrence or reflux after treatment. Such a system would not replace the endoscopist but could create a more consistent and data-driven care pathway. Figure 1 summarizes this proposed workflow, while Table 1 summarizes the evidence status of the major applications.

Safety, Generalizability, and Governance

Dataset shift is a central safety problem. A model trained in a high-volume expert center may fail when case mix, disease prevalence, endoscope manufacturer, processor, image enhancement, catheter, FLIP protocol, operator technique, or annotation practice changes. Class imbalance may conceal poor sensitivity for uncommon but consequential events such as mucosal injury. Validation should therefore include geographically and technically distinct centers, temporal testing, calibration plots, and prespecified subgroup performance by sex, age, ethnicity, disease subtype, previous treatment, and procedural complexity.

Fairness cannot be inferred from a single overall accuracy. Underrepresented groups may receive less reliable estimates, and missing data may encode unequal access to specialist testing. Explainability should be matched to the clinical task. False alerts can distract the operator, whereas false reassurance can delay recognition of a complication. Automation bias, alert fatigue, and the effect of AI on trainee skill acquisition require direct human-factors evaluation.

Governance must address lawful data use, privacy protection for endoscopy video and physiological data, cybersecurity, version control, audit logs, regulatory responsibility, and clear assignment of accountability when the clinician and system disagree. Institutions should define who may override an output, how failures are reported, and how performance is monitored after deployment. Prospective surveillance should assess drift, near misses, subgroup harm, and workflow effects, with a predefined withdrawal or recalibration process.

Prediction-model reports should follow the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis plus AI statement, and risk of bias should be assessed with the prediction model risk of bias assessment tool plus AI (33,34). Early live clinical studies should address the developmental and exploratory clinical investigations of decision-support systems driven by AI guidance, while the fairness, universality, traceability, usability, robustness, and explainability for AI framework informs the wider lifecycle (35,36). Table 1 summarizes the evidence maturity, validation status, and principal limitation of each application.

Novelty of This Review

The novelty of this narrative review is its stage-specific evidence-maturity framework across the complete POEM pathway. It distinguishes direct achalasia or POEM evidence from adjacent therapeutic-endoscopy research and conceptual proposals, and it

links each application to its validation status, principal limitation, and required clinical safeguards. This approach provides a clinically oriented research agenda without implying that conceptual procedural tools are ready for patient care.

CONCLUSION

AI-specific applications across the POEM pathway remain at an early stage, although conventional regression models provide internally validated precursor evidence for poor response, technical difficulty, and major adverse events. The most realistic near-term AI uses are clinician-facing detection support during diagnostic endoscopy, bounded assistance with HRM or FLIP interpretation, retrospective video review, and cautiously calibrated outcome stratification. Even these uses require external and prospective validation in representative populations before routine adoption.

Personalized myotomy prescription, real-time tunnel navigation, vessel or mucosal-injury alerts, and automated training assessment remain speculative for POEM. Progress requires multicenter datasets, standardized outcomes, calibration and subgroup analyses, human-factors and clinical-impact studies, regulatory clarity, and post-deployment monitoring. AI should augment expert judgment and must not weaken established diagnostic, procedural, or reflux follow-up safeguards.

Acknowledgments

Artificial intelligence was used only for language editing and grammatical corrections. The authors reviewed and approved the final manuscript and take full responsibility for its content.

Footnotes

Authorship Contributions

Concept: A.A.S., Design: A.A.S., A.A., Data Collection or Processing: A.A.S., A.A., Analysis or Interpretation: A.A.S., A.A., Literature Search: A.AS., A.A., Writing: A.A.S., A.A.

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

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

REFERENCES

1. Inoue H, Sato H, Ikeda H, Onimaru M, Sato C, Minami H, et al. Per-oral endoscopic myotomy: a series of 500 patients. J Am Coll Surg. 2015; 221: 256-64.
2. Calabrese EC, Kindel T, Slater BJ, Marks J, Swanstrom L, Saxena P, et al. 2024 update to SAGES guidelines for the use of peroral endoscopic myotomy (POEM) in the treatment of achalasia. Surg Endosc. 2025; 39: 4027-37.
3. Kohn GP, Dirks RC, Ansari MT, Clay J, Dunst CM, Lundell L, et al. SAGES guidelines for the use of peroral endoscopic myotomy (POEM) for the treatment of achalasia. Surg Endosc. 2021; 35: 1931-48.
4. Werner YB, Hakanson B, Martinek J, Repici A, von Rahden BHA, Bredenoord AJ, et al. Endoscopic or surgical myotomy in patients with idiopathic achalasia. N Engl J Med. 2019; 381: 2219-29.
5. Bencardino S, Lodola I, Centanni L, Mandarino FV, Fanizza J, Furfaro F, et al. Artificial intelligence in advanced endoscopic imaging: transforming optical diagnosis in gastroenterology. Front Med (Lausanne). 2026; 12: 1719145.
6. Kaklamanos E, Kristinsdottir K, Wittbrodt M, Li M, Pitisuttithum P, Kou W, et al. From pixels to peristalsis: comparing achalasia diagnosis with artificial intelligence with expert endoscopists. Clin Gastroenterol Hepatol. 2026; 24: 1542-9.
7. Shiwaku H, Misawa M, Inoue H, Jiang K, Oda M, Familiari P, et al. Computer-aided detection for esophageal achalasia (with video). Dig Endosc. 2025; 37: 851-61.
8. Yadlapati R, Kahrilas PJ, Fox MR, Bredenoord AJ, Prakash Gyawali C, Roman S, et al. Esophageal motility disorders on high-resolution manometry: Chicago classification version 4.0©. Neurogastroenterol Motil. 2021; 33: e14058. Erratum in: Neurogastroenterol Motil. 2024; 36: e14179.
9. Fass O, Rogers BD, Gyawali CP. Artificial intelligence tools for improving manometric diagnosis of esophageal dysmotility. Curr Gastroenterol Rep. 2024; 26: 115-23.
10. Czako Z, Surdea-Blaga T, Sebestyen G, Hangan A, Dumitrascu DL, David L, et al. Integrated relaxation pressure classification and probe positioning failure detection in high-resolution esophageal manometry using machine learning. Sensors (Basel). 2021; 22: 253.
11. Carlson DA, Kou W, Rooney KP, Baumann AJ, Donnan E, Triggs JR, et al. Achalasia subtypes can be identified with functional luminal imaging probe (FLIP) panometry using a supervised machine learning process. Neurogastroenterol Motil. 2021; 33: e13932.
12. Urakami S, Abe H, Tanaka S, Kawara F, Toyonaga T, Ariyoshi R, et al. Development of a preoperative risk-scoring system for predicting poor responders to peroral endoscopic myotomy. Gastrointest Endosc. 2021; 93: 398-405.
13. Liu X, Yao L, Cheng J, Xu M, Chen S, Zhong Y, et al. Landscape of adverse events related to peroral endoscopic myotomy in 3135 patients and a risk-scoring system to predict major adverse events. Clin Gastroenterol Hepatol. 2021; 19: 1959-66.e3.
14. Liu XY, Geng ZH, Chen WF, Xu MD, Chen SY, Zhong YS, et al. A prediction model and nomogram for technical difficulty of peroral endoscopic myotomy. Surg Endosc. 2023; 37: 2781-8.
15. Takahashi K, Sato H, Shimamura Y, Abe H, Shiwaku H, Shiota J, et al. Achalasia phenotypes and prediction of peroral endoscopic myotomy outcomes using machine learning. Dig Endosc. 2024; 36: 789-800.
16. Khashab MA, Vela MF, Thosani N, Agrawal D, Buxbaum JL, Abbas Fehmi SM, et al. ASGE guideline on the management of achalasia. Gastrointest Endosc. 2020; 91: 213-27.e6.
17. Rohof WO, Hirsch DP, Kessing BF, Boeckxstaens GE. Efficacy of treatment for patients with achalasia depends on the distensibility of the esophagogastric junction. Gastroenterology. 2012; 143: 328-35.
18. Hasan A, Low EE, Fehmi SA, Yadlapati R. Evolution and evidence-based adaptations in techniques for peroral endoscopic myotomy for achalasia. Gastrointest Endosc. 2022; 96: 189-96.
19. Srinivasan V, Shah ED. Mini-review: tailored per-oral endoscopic myotomy for type III achalasia. Neurogastroenterol Motil. 2023; 35: e14700.
20. Vespa E, Barchi A, Mandarino FV, Fasulo E, Fratto MC, Passaretti S, et al. Standard length of peroral endoscopic myotomy (POEM) for achalasia: a systematic review and meta-analysis. Dis Esophagus. 2024; 37: doae069.
21. Knowles TB, Jackson AS, Chang SC, Schembre DB, Farivar AS, Aye RW, et al. Changes in distensibility index during an incremental POEM myotomy. J Gastrointest Surg. 2022; 26: 1140-6.
22. Knight W, Kandiah K, Vrakopoulou Z, White A, Barbieri L, Tewari N, et al. Early outcomes following EndoFLIP-tailored peroral endoscopic myotomy (POEM): the establishment of POEM services in two UK centers. Dis Esophagus. 2023; 36: doac110.

23. Kim MJ, Jeon SY, Park JC, Youn YH, Park H. Loss of inspiratory augmentation as a predictive factor for the development of gastroesophageal reflux disease after peroral endoscopic myotomy in patients with achalasia. J Neurogastroenterol Motil. 2026; 32: 129-35.
24. Cao J, Yip HC, Chen Y, Scheppach M, Luo X, Yang H, et al. Intelligent surgical workflow recognition for endoscopic submucosal dissection with real-time animal study. Nat Commun. 2023; 14: 6676.
25. Scheppach MW, Yip HC, Chen Y, Yang H, Cao J, Chua T, et al. Feasibility of real-time artificial intelligence-assisted anatomical structure recognition during endoscopic submucosal dissection. Endosc Int Open. 2025; 13: a26158008.
26. Zhu Y, Du L, Fu PY, Geng ZH, Zhang DF, Chen WF, et al. An automated video analysis system for retrospective assessment and real-time monitoring of endoscopic procedures (with video). Bioengineering (Basel). 2024; 11: 445.
27. Arakaki S, Takenaka S, Sasaki K, Kitaguchi D, Hasegawa H, Takeshita N, et al. Artificial intelligence in minimally invasive surgery: current state and future challenges. JMA J. 2025; 8: 86-90.
28. Nabi Z, Reddy DN, Ramchandani M. Adverse events during and after per-oral endoscopic myotomy: prevention, diagnosis, and management. Gastrointest Endosc. 2018; 87: 4-17.
29. Dacha S, Aihara H, Anand GS, Byrne KR, Chahal P, James T, et al. Core curriculum for peroral endoscopic myotomy (POEM). Gastrointest Endosc. 2021; 93: 539-43.
30. Yang D, Draganov PV, Pohl H, Aihara H, Jeyalingam T, Khashab M, et al. Development and initial validation of a video-based peroral endoscopic myotomy assessment tool. Gastrointest Endosc. 2024; 99: 177-85.
31. Repici A, Fuccio L, Maselli R, Mazza F, Correale L, Mandolesi D, et al. GERD after per-oral endoscopic myotomy as compared with Heller's myotomy with fundoplication: a systematic review with meta-analysis. Gastrointest Endosc. 2018; 87: 934-43.e18.
32. Inoue H, Ueno A, Shimamura Y, Manolakis A, Sharma A, Kono S, et al. Peroral endoscopic myotomy and fundoplication: a novel NOTES procedure. Endoscopy. 2019; 51: 161-4.
33. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024; 385: e078378. Erratum in: BMJ. 2024; 385: q902.
34. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025; 388: e082505.
35. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al.; DECIDE-AI expert group. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ. 2022; 377: e070904.
36. Lekadir K, Frangi AF, Porras AR, Glocker B, Cintas C, Langlotz CP, et al.; FUTURE-AI Consortium. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ. 2025; 388: e081554. Erratum in: BMJ. 2025; 388: r340.